RICE UNIVERSITY

Structural and Functional Studies of the human cohesin subunits Rad21 and SA2

by

Yunyun Jiang

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Yizhi Jane Tao, Ph.D., Associate Professor, Advisor
Biochemistry and Cell Biology

Debananda Pati, Ph.D., Associate Professor, Advisor
Department of Pediatrics, Section of Hematology-Oncology,
Baylor College of Medicine

James A. McNew, Ph.D., Associate Professor,
Biochemistry and Cell Biology

Jonathan Silberg, Ph.D., Assistant Professor,
Biochemistry and Cell Biology

Weiwei Zhong, Ph.D., Assistant Professor,
Biochemistry and Cell Biology

Robert M. Raphael, Ph.D., Associate Professor,
Bioengineering

HOUSTON,TEXAS
JULY 2011

Abstract

Structural and functional analysis of the human cohesin subunits Rad21 and SA2

Yunyun Jiang

The cohesin complex is responsible for the fidelity of chromosomal segregation during mitosis. It consists of four core subunits namely Rad21/Mcd1/Scc1, Smc1, Smc3 and one of the yeast Scc3 orthologs SA1 or SA2. Sister chromatid cohesion is formed by the cohesin complex during DNA replication and maintained until the onset of anaphase. Among the many proposed models of how cohesin holds sister chromatids together, the 'core' cohesin subunits Smc1, Smc3 and Rad21/Mcd1/Scc1 are almost universally displayed as forming a contiguous ring. However, other than its supportive role in the cohesin ring, little is known about the fourth core protein SA1/SA2 - despite its physical association to the cohesin ring. To gain deeper insight into how physically and physiologically SA2 interacts with the cohesin complex, we performed structural characterization of SA2 and Rad21 and mapped the interaction region of the two proteins *in vitro* and *ex vivo*. We found SA2 interacts with Rad21 at multiple domains while Rad21 only interacts with SA2 through a 10 amino acid α-helical motif from 383-392aa. Deletion of these 10 amino acids or mutation of three conserved amino acids (L385, F389, and T390) in this α-helical motif prevents Rad21 from physically interacting with SA2/SA1 and causes premature sister chromatid separation in mitotic cells that often leads to aneuploidy. Our studies provide a model for how SA2 structurally strengthens the cohesin ring through its interaction with Rad21. Results from our structural characterization of these two proteins also provided directions for further investigation of the structural basis of protein-protein interaction in the cohesin complex.

Acknowledgements

I would like to thank my advisor Dr. Yizhi Jane Tao for her support throughout my years in graduate school. Dr. Tao serves an excellent model for me and for everyone in the lab by showing her great passion in science. She has created a positive atmosphere and a high standard for me and others to perform research in the lab. She not only helped me with the scientific techniques, but also encouraged me to think thoroughly and critically. She is not only an excellent scientist, but also a great mentor. By supporting me to attend keystone conferences, she provided me the opportunity to recognize the work of many excellent fellow scientists and keep up with the newest findings in the field. As a result of Dr. Tao's mentorship, I have been able to deal with the difficulties in research and continuously improve myself.

I would like to thank Dr. Debananda Pati, my co-advisor, for his help in my research. I appreciate his efforts in helping me by focusing on scientific issues. I am very grateful and honored to have him as my co-advisor. I would like to thank him for his expertise in cohesin and nurturing environment he provided in his lab for me to perform part of the experiments for my thesis.

I am grateful to my committee, Dr. James McNew, Dr. Jonathan Silberg, Dr. Weiwei Zhong, Dr. Robert Raphael and former committee member Dr. Kevin McKenzie. I appreciate the efforts they made in creating a flexible environment for my research. I would like to thank them for their scientific suggestions and their encouragement, which helped me to improve continuously.

Our collaborator, Dr. Borries Demeler (University of Health Science Center, San Antonio, TX) graciously provided training in the Ultrascan software. I would also like to thank him for his help in data analysis.

I would like to thank the current and former members of the Tao laboratory, Dr. Junhua Pan, Dr. Qiaozhen Ye, Dr. Tom Guu, Ms. Liping Dong, Mr. Zanhua Jack Xie, Dr. Li Lin, Dr. Jinhui Dong, Dr. Yukimatsu Toh, Mr. Wenjie Zheng, Mr. Aaron Collier, Ms. Yangyang Dong and Ms. Yusong Guo, for creating a pleasant and harmonious atmosphere for the lab. I would like to thank Dr. Pan and Dr. Ye for sharing their experience in protein purification. Dr. Ye and Dr. Guu guided me through the insect cell expression techniques. Mr. Xie provided great help in electron microscopy sample preparation with both negative staining and metal shadowing methods. I am very grateful to Ms. Liping Dong not only for her help with specific lab techniques, but also for her encouragement when I faced frustration.

I would like to thank the members of Pati Laboratory, Dr. Nenggang Zhang, Ms. Guoqing (Mary) Ge, Dr. Anil Panigrahi, Dr. Elizabeth McCullum, Dr. Qilong Mao and Dr. Malini Mukherjee. Dr. Zhang provided a great deal of help in the mammalian cell related experiments and he also provided valuable training in mammalian cell culture techniques.

I would like to thank Dr. Susan Cates for providing valuable training in ultracentrifugation training. I would like to thank the laboratories of Dr. Yousif Shamoo,

Dr. Kevin McKenzie, Dr. John Olsen and Dr. Kathleen Matthew for sharing instruments and various supplies. I would like to thank Dr. Wenhua Guo for providing training in TEM techniques.

I appreciate the wonderful friendship with Dr. Ying Frances Liu, Ms. Liping Dong, Dr. Cassidy Johnson and Ms. Sol Gomez de la Torre Canny and many others. I would to thank them for their cordial help in all matters.

I appreciate all the support and help provided through the years by the Department of Biochemistry and Cell Biology, the Office of International Students and Scholars, the Rice Chinese Student and Scholar's Club and Rice University. I would like to thank Ms. Dolores Schwartz for her generous help when I first came to the United States.

I would like to thank my parents for giving me life and supporting me through the years. Nothing would have been possible without their unconditional love and everything they have done for me.

I apologize if I have unintentionally left anyone's name out.

Table of Contents

List of Figures

List of Tables

List of Terms, Symbols and Abbreviation

APC/C	anaphase-promoting complex or cyclosome
aa	amino acid
Amp	ampicillin
AU	analytical ultracentrifugation
CD	circular dichroism
cDNA	complementary DNA
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EV	empty vector
FPLC	fast protein liquid chromatography
Gen	gentamicin
G(s)	Gaussian distributions of the sedimentation coefficient
GSSG	glutathione
GSH	glutathione disulfide
His	histidine
HPLC	high performance liquid chromatography
IP	immunoprecipitation
IPTG	isopropyl β-D-l-thiogalactopyranoside
Kan	Kanamycin
LB	lysogeny broth

mAb	monoclonal antibody
MALDI-MS	matrix-assisted laser desorption/ionization mass spectroscopy
MOI	multiplicity of infection
MQ	Milli-Q
MW	molecular weight
NaCl	sodium chloride
NBD	nuclear binding domain
Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
OD	optical density
PA	acidic protein subunit of the influenza A virus polymerase
PBS	phosphate buffered saline
PCR	olymerase chain reaction
PEG	polyethylene glycol
PMSF	phenymethylsulphonyl fluoride
PVDF	polyvinylidene fluoride
pAb	poly clonal antibody
pI	isoelectric point
RNA	ribonucleic acid
SDS-PAGE	sodium dedecylsulfate polyacrylamide gel electrophoresis
SE	sedimentation equilibrium
SMC	structure maintenance of chromosomes

siRNA	small interfering RNA
SV	sedimentation velocity
TBS	Tris buffered saline
TEM	transmitted electron microscopy
Tet	tetracycline
Tris	tris (hydroxymethyl) aminomethane
UTR	upstream
WHD	winged helix domain
WT	wildtype
YFP	yellow fluorescent protein

Chapter I. Introduction

1.1 Cancer and aneuploidy

Aneuploidy is a type of chromosome abnormality, in which the number of total chromosomes is not an exact multiple of a haploid set. In human, aneuploidy is a characteristic feature of most if not all cancers (reviewed by Panigrahi *et al.*, 2009). Previous data has shown that aneuploidy is frequently observed in solid tumors including breast cancer and non-small cell lung cancer tissues (Choma *et al.*, 2001). Aneuploid tumors have been found to be correlated with the poorer prognosis of breast, gastric, pancreatic cancer comparing to the diploid tumors (Nakopoulou *et al.*, 2007; Ikeguchi *et al.*, 1995; Sciallero *et al.*, 1993). Therefore, a better understanding of the cause of aneuploidy leads to better understanding of tumorigenesis and the development of therapeutic methods.

Aneuploidy can be a result of inaccurate chromosomal segregation. Chromosomal missegregation can be caused by many factors, one of which is the defects in sister chromatid cohesion and separation. The cohesion between sisters chromatids forms during the replication of chromosomes in S phase. In metaphase, the sister chromatids are pulled by microtubules through the kinetochores towards opposing spindle poles (Nasmyth, 2001). If there is no resistance against the pulling force, the sister chromatids will likely be pulled towards the spindle poles in an uncontrolled manner. The cohesion between sister chromatids provides the resistance against the pulling. When the resistance and the pulling force reach equilibrium, sister chromatids are able to align in the same plane. Then, the cohesion between sister chromatids is destroyed and sister chromatids

are separated from each other, completing the transition from metaphase to anaphase. The failure in the formation and maintenance of the cohesion between sister chromatids results in premature sister chromatid segregation, which is thought to be a major pathway to aneuploidy. On the other hand, inadequate separation of the sister chromatids can result in the lagging of chromosomal segregation to the opposite poles, thus culminating in aneuploidy.

1.2 Introduction to the cohesin complex

1.2.1 Overview of the cohesin complex and its subunits

The chromosomal cohesion (i.e., holding the sister chromatids together) is provided by a multi-subunit complex called cohesin. The mitotic cohesin complex is conserved from yeast to human (Guacci *et al.*, 1997; Michaelis *et al.*, 1997; Darwiche *et al.*, 1999; Vass *et al.*, 2003). It consists of four core subunits, which are known as Rad21 (Mcd1/Scc1 in yeast), Structure Maintenance of Chromosomes (SMC) proteins, Smc1 and Smc3, and either one of the yeast Scc3 orthologs, SA1 or SA2 (also known as stromal antigens STAG1 and STAG2) (Uhlmann, 2001; Hagstrom *et al.*, 2003; Losada *et al.*, 2000).

In addition to the 'core' subunits, there are other subunits known to associate with the cohesin complex, such as Pds5 (Pds5a and Pds5b in vertebrates), Wapl (Rad61 in yeast) (Panizza *et al.*, 2000; Losada *et al.*, 2005; Kueng *et al.*, 2006; Nishiyama *et al.*, 2010) and Sororin (Schmitz *et al.*, 2007).

1.2.2 Introduction to the cohesin subunits

Since the identification of the cohesin core subunit more than a decade ago (Guacci *et al.*, 1997), intensive research has been conducted to functionally characterize the cohesin subunits. Table 1 lists the main cohesin subunits in yeast and human and their general information, including their size and sequence homology between organisms.

Cohesin subunits		Number of aa/MW		Sequence similarity between the yeast and human protein
Yeast	Human	Yeast	Human	
Scc1/ Mcd1	Rad21	566/63.2kD*	631/71.7kD*	~30%
Smc1	Smc1	1225/141kD	1233/143kD	~54%
Smc3	Smc3	1230/141kD	1217/142kD	~56%
Scc3	SA1/2	1150/133kD	SA1:1258/141kD SA2: 1231/141kD	Yeast Scc3 and SA1: 38% Yeast Scc3 and SA2: 31%
Pds5	Pds5a, Pds5b	1150/133kD	Pds5a: 1337/151kD Pds5b: 1447/165kD	Yeast Pds5 and Human Pds5a/Pds5b: 41%
Rad61	Wapl	647/74.7kD	1190/133kD*	~20%

Table 1. A list of the information of human cohesin complex subunits from the NCBI database. * : Both human and yeast Rad21 show an apparent MW of ~120kD in an SDS-PAGE gel and the human Wapl shows an apparent MW of ~170kD in an SDS-PAGE gel.

SMC proteins are large coiled coil proteins which form the 'backbone' of the cohesin complex. They are consisted of five domains, in which three of them are globular and the other two are 50nm-long coiled coil regions (Figure 1.1a; Melby *et al.*, 1998).

The two long coiled coil regions are formed in an anti-parallel manner, connecting the three globular domains, two of which are encoded by amino acids from the N- and the C-termini of the protein respectively, and the third one is encoded by amino acids from the center of the protein. Rotary metal shadowed electron-micrographs have revealed that in the cohesin complex, the center globular region of the SMC proteins each forms a hinge domain, via which, Smc1 and Smc3 form a V shaped heterodimer (Anderson *et al.*, 2002; Haering *et al.*, 2002). Co-purification of different domains of the yeast cohesin SMC proteins from insect cells also confirmed that only the hinge domains, but not other domains, of the SMC proteins could co-purify with each other (Haering *et al.*, 2002). This finding is consistent with the conclusion from the EM observation. The crystal structure of the mouse hinge domain has recently been solved (Figure 1.1b; Kurze *et al.*, 2010), suggesting that the hinge domain contains a positively charged channel, which potentially functions as an entry gate for positively charged DNA molecules. On the other end of the coiled coil region, the N- and C-terminus of the SMC protein form an ATP binding cassette (ABC) like nucleotide-binding domain (NBD) (Haering *et al.*, 2002; Hirano *et al.*, 2002).

Previous studies have found that the 115 amino acids (aa) from the N-terminus of the yeast Rad21 could co-purify with the Smc3 NBD and the C-terminal 180aa of Rad21 could co-purify with the Smc1 NBD (Haering *et al.*, 2002; Gruber *et al.*, 2003), suggesting Rad21 binds to Smc3 and Smc1 via its N- and C-terminus, respectively. When Rad21 is cleaved by Separase, the N- and the C-terminus of Rad21 still remain associated with each other, suggesting that Rad21, Smc1 and Smc3 form a ring (Gruber *et al.*, 2003).

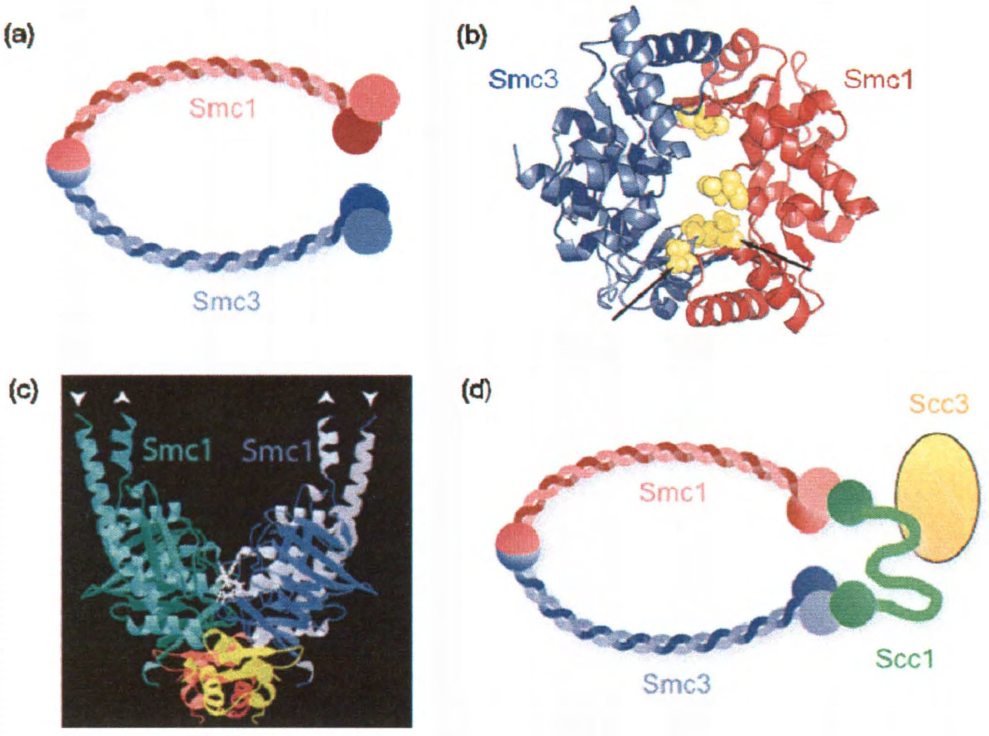


Figure 1.1 Structures of the cohesin subunits. (a) Illustration of the cohesion SMC protein structures (adapted from Haering *et al.*, 2002). (b) Crystal structure of the mouse Smc3-Smc1 hinge domain (from Kurze *et al.*, 2010), the residues marked in yellow form a positively charged channel. (c) Crystal structure of the yeast Scc1 C-terminus in complex with the yeast Smc1 ATPase head domain (from Haering *et al.*, 2004). (d) In yeast, Scc3 associates with the cohesin complex through binding to Scc1 (adapted from Haering *et al.*, 2002).

The structure of the yeast complex formed by the C-terminus of Rad21 and the Smc1 NBD has been solved (Figure 1.1c; Haering *et al.*, 2004). The C-terminus of Rad21 forms a winged-helix domain (WHD) which binds to two C-terminal β-strands of the Smc1 NBD through hydrophobic interactions. The complex contains an ATP binding pocket. The ATP hydrolysis catalyzed by the ATPase domain (i.e., NBD) is required for cohesin's association with chromosomes (Arumugam *et al.*, 2003). The WHD increased the ATPase activity of Smc1/Smc3 by increasing the binding of ATP to the Smc1 head (Haering *et al.*, 2004; Arumugam *et al.*, 2006). The N-terminus of Rad21 and the Smc3 head domain also forms an ATPase domain (Haering *et al.*, 2003). Interestingly, mutation of amino acids either within the Rad21 C-terminal WHD or the Smc1 C-terminal β-strands eliminates all association between Scc1 and Smc1/*3 in vivo* (Haering *et al.*, 2004), which suggests that the WHD-Smc1 interaction may initiate the association of Scc1 to the Smc1/Smc3 heterodimer, after which the binding between the Smc1 NBD and the N-terminus of Scc1 takes place.

It has been reported that the yeast Scc3 does not directly bind to the Smc1/Smc3 heterodimer (Figure 1.1d; Haering *et al.*, 2002). It associates with the cohesin ring through direct binding to the C-terminal Separase cleavage product of Rad21 (269-566aa) (Haering *et al.*, 2002; Gruber *et al.*, 2003). However, it is unknown how the human SA1/2 associates with the cohesin complex. The two human Scc3 orthologs, SA1 and SA2, do not co-exist in the same cohesin complex (Losada *et al.*, 2000). The relative abundance of the two orthologs varies among organisms. For example, Xenopus egg extracts contains ten times more cohesin^{SA1} than cohesin^{SA2}, while HeLa cell extracts contains three times more cohesin^{SA2} than cohesin^{SA1} (Losada *et al.*, 2000). The

biological significance of the variation in the relative abundance of the two Scc3 orthologs among different organisms is unknown.

In humans, the SA1 and SA2 share 70% overall sequence identities (Figure 1.2). For SA2, most of the amino acid variations with SA1 are located in the N-terminal 1-68aa and the C-terminal 1075-1162aa regions. SA1 and SA2 have been reported to play different roles in chromosomal cohesion. SA1 has been shown to preferentially involve in telomere cohesion, which also affects arm cohesion (Canudas *et al.*, 2009). SA2 might play a role in arm cohesion and centrimeric cohesion (Hauf *et al.*, 2005). It is not known how the cohesin^{SA1} and cohesin^{SA2} can function differently in the sister chromatid cohesion. It is possible that they are located at different loci on the chromosome to perform their different biological function.

Pds5 (precocious dissociation of sister chromatids) is composed of HEAT repeats and there are two Pds5 orthologs Pds5A and Pds5B in vertebrates (Losada *et al.*, 2005). Compared to the cohesin 'core' subunits, it binds more loosely to the cohesin complex and it maintains sister chromatid cohesion during G2 phase (Hartman *et al.*, 2000; Panizza *et al.*, 2001). It is found that Pds5 localized to the chromatin in a cell-cycle dependent manner similar to Rad21 (Hartman *et al.*, 2000). In yeast, Pds5 associates with the cohesin complex through binding to Scc3 (Rowland *et al.*, 2009). However, this binding only occurs in the existence of the yeast Wapl homolog, Rad61 and the three proteins are able to form a Scc3-Rad61-Pds5 complex *in vitro* (Rowland *et al.*, 2009). In humans, it was found that, Pds5B has the intrinsic property of binding with Rad21. However, the binding is stronger in the presence of SA1 (Shintomi *et al.*, 2009). It is not known if Pds5A binds to Rad21 and SA1 in the same manner.

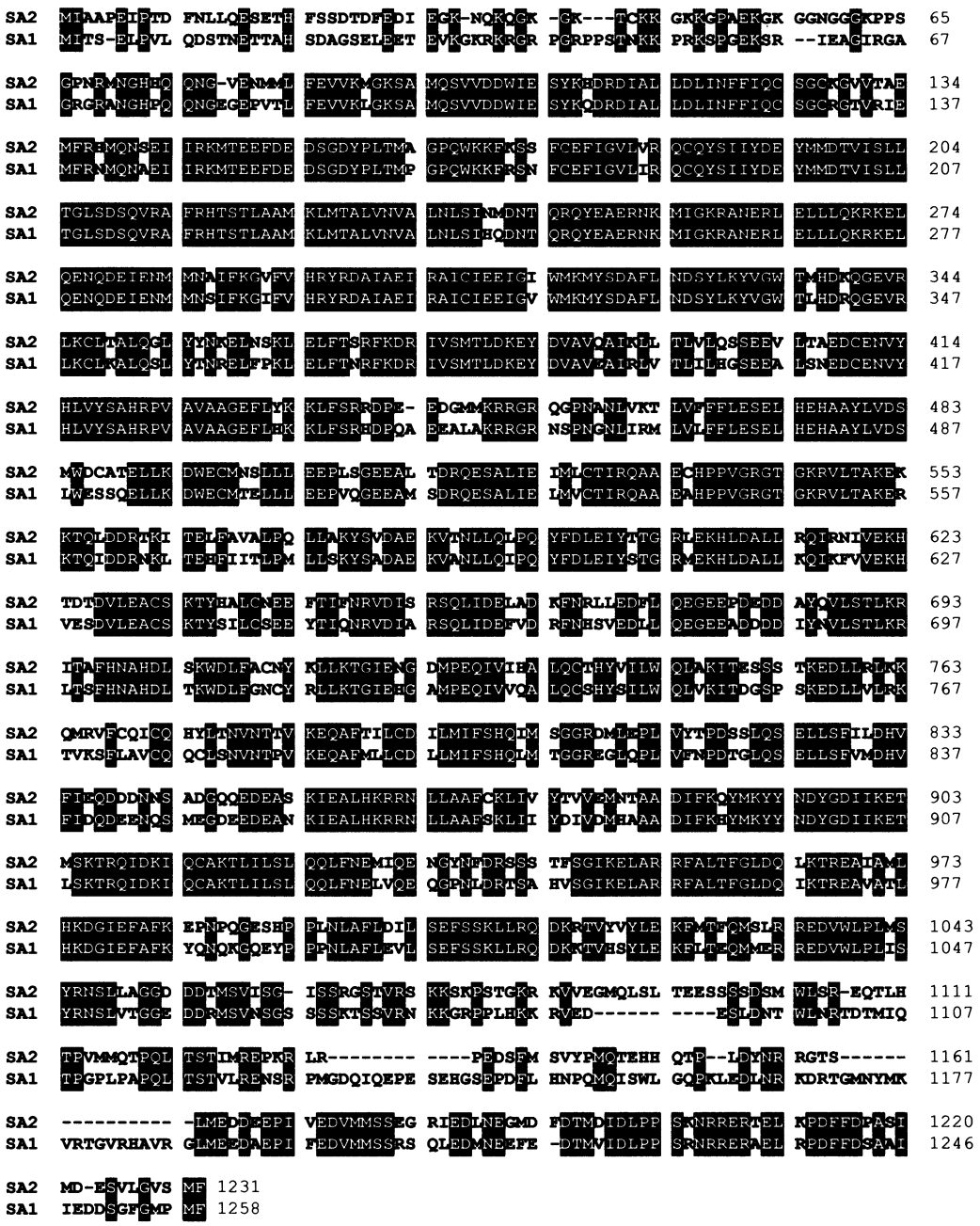


Figure 1.2 Sequence alignment of the human SA1 and SA2. The sequence alignment is prepared using BioEdit (Hall, 1999).

Wapl forms stable complex with Pds5A in HeLa cells (Kueng *et al.*, 2006). The depletion of Wapl increases the number of cohesin complexes associated with chromatin and prolonged the cohesin association during mitosis. It is thus likely that Wapl plays a role in the cohesin dissociation possibly by having an indirect effect on the opening of the cohesin ring either through the ATPase domain or the Smc1/Smc3 hinge domain.

Sororin, as a Wapl antagonist, has only been identified in vertebrates (Rankin *et al.*, 2005; Nishiyama *et al.*, 2010). Sororin maintains the cohesion of sister chromatids by inhibiting the dissociation of cohesin from chromosomes caused by Wapl (Nishiyama *et al.*, 2010).

1.2.3 The models of the cohesin complex

There are several models proposed to explain how the cohesin complex holds the sister chromatids together (Anderson *et al.*, 2002; Haering, *et al.*, 2003; Huang *et al.*, 2006; Milutinovich *et al.*, 2003; Zhang *et al.*, 2008). Among them, the most dominant ones are the one-ring embrace model (Gruber *et al.*, 2003; Haering *et al.*, 2003) and the two-ring handcuff model (Zhang *et al.*, 2008).

The one-ring embrace model (Figure 1.3 a) was proposed in yeast. It has already been shown that Smc1, Smc3 and Scc1 form a tripartite ring, while Scc3 associates with the ring through binding to Scc1 (Haering *et al.*, 2002; Gruber *et al.*, 2003). In one-ring model, both of the sister chromatids are entrapped in the same cohesin ring. At the onset of anaphase, Separase cleaves Rad21, which opens the ring and releases sister chromatids from cohesion. This model explains how the 38nm cohesin ring holds yeast sister

chromatids each with a 9nm diameter together. However, when it comes to higher eukaryotes, the diameter of each sister chromatid is 30nm, and therefore the ring is not big enough to entrap both sister chromatids. The region of sister chromatids near the cohesin binding sites may undergo some topological change to allow the loading of the cohesin complex. However, it is now known how this exactly takes place.

To solve the contradiction raised from the size, a handcuff model was proposed (Zhang *et al.*, 2008). In the handcuff model, there are two sets of Smc1, Smc3 and Rad21, each forming a cohesin ring and each ring entraps one sister chromatid. This model was proposed based on the finding that differently tagged Smc1, Smc3 and Rad21 can pull down each other in mammalian cells. However, neither SA1 nor SA2 was able to pull down itself, suggesting that there is only one molecule of either SA1 or SA2 in each cohesin complex. A Fluorescent protein fragment complementary assay showed that two yellow fluorescent proteins (YFP) fused Rad21 molecules can only emit fluorescence when the YFP tags are located at different termini of the Rad21 molecules, indicating that the two Rad21 molecules interact with each other in an anti-parallel manner. The Rad21-Rad21 interaction is significantly reduced when SA1/SA2 is knocked down from the cells. Therefore, the interaction is dependent on SA1/SA2. The handcuff model suggests that the two rings are held together by the anti-parallel interaction of the two Rad21 subunits and SA1/SA2 fortifies the handcuff (Figure 1.3b). The handcuff model explains the contradiction from the one-ring embrace model, in which one ring is not big enough to hold sister chromatids. However, it is unclear if the dimerization of Rad21 is directly caused by the SA1/SA2 subunit or through other cohesin associated components.

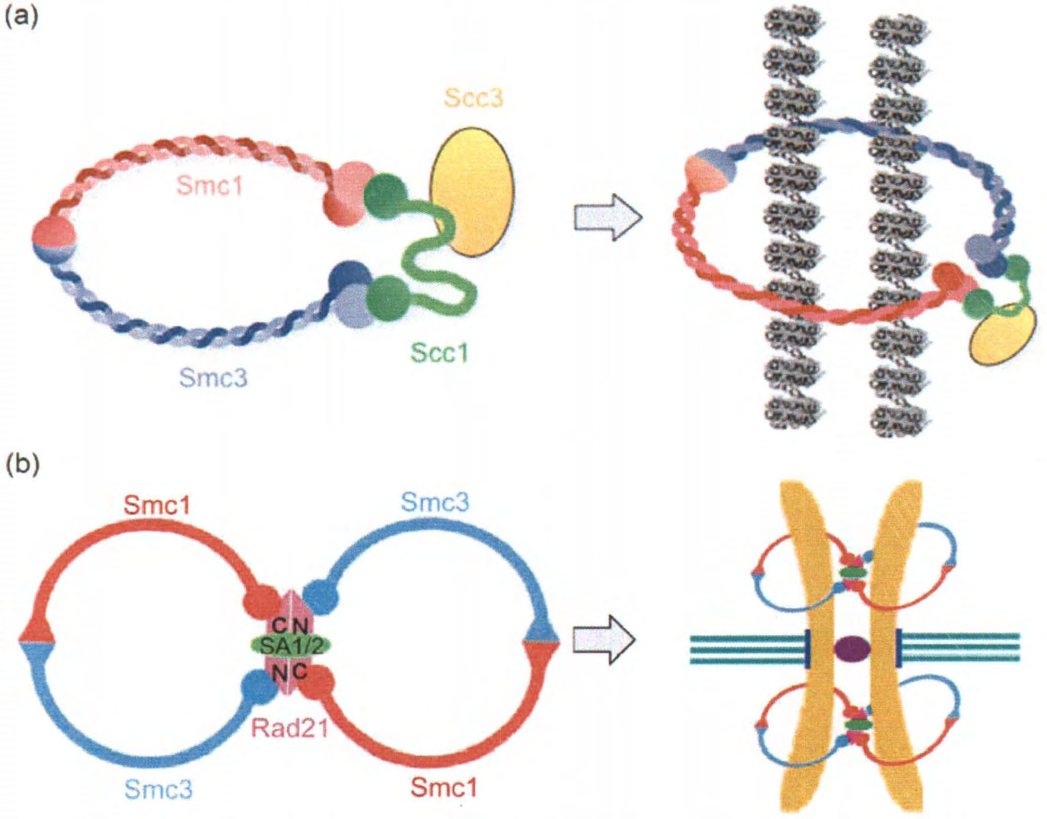


Figure 1.3 The models of the cohesin complex. (a) One-ring embrace model from yeast (Adapted from Haering *et al.*, 2002). (b) Two-ring handcuff model from humans (from Zhang *et al.*, 2008).

1.2.4 The establishment of the cohesion

The cohesin complex associates with chromosomes before DNA replication (Hauf *et al.*, 2005). The cohesin complexes are loaded onto chromosomes via the opening of the Smc1/Smc3 hinge by the Scc2-Scc4 loading complex (Ciosk *et al.*, 2000; Gruber *et al.*, 2006). Scc2 contains multiple HEAT repeats and Scc4 consists of TPR repeats. The Scc2-Scc4 complex is highly conserved from yeast to human. It is unclear how exactly the Scc2-Scc4 complex promotes the cohesin loading. It is likely that the complex helps to open the cohesin ring by facilitating the changes in topology of the ring and/or bringing the cohesin complex closer to chromosomes for them to associate. The Scc2-Scc4 complex may also be able to modify the chromatid fiber to ease its entry into the cohesin ring. Recent studies have also shown Esco1 (Eco1 in yeast), which is an acetyl transferase, facilitates the establishment of cohesion by modifying Smc3 at two consecutive lysine sites in its ATPase domain (Terret *et al.*, 2009; Peters *et al.*, 2009). The Smc3 acetylation seems to increase in S phase, which is suggested to depend on the Scc2-Scc4 loading complex (Unal *et al.*, 2008). The acetylation starts decreasing at anaphase which is suggested to be triggered by the dissociation of the cohesin complex from sister chromatids (Nasmyth *et al.*, 2009).

It was mentioned in section 1.2.2 that cohesin^{SA1} and cohesin^{SA2} might localize differently on the chromatin which resulted in their different functions. It is unknown how cohesin^{SA1} and cohesin^{SA2} in vertebrates may be loaded to different loci of the chromatin as to function differently. SA1 and SA2, the Scc3 subunits, share high sequence identity in most part but vary in both N- and C- termini. The difference in their sequences can lead to their binding to different proteins. The different proteins they bind

to might cause them to locate on different loci on the chromosomes. For example, SA1 is found to regulate telomere cohesion. The telomeric protein TIN2 may assist in the localization of cohesin^{SA1} for the cohesion regulation to take place. Another possibility is that the different sequences lead to different structures for both proteins, especially at both N- and C-termini, thus affects the overall cohesin complex confirmation. The difference in the conformation of different cohesin complexes can lead to their selective binding to different loci of the chromatin. It is not known whether the Scc2/Scc4 complex is involved in this process.

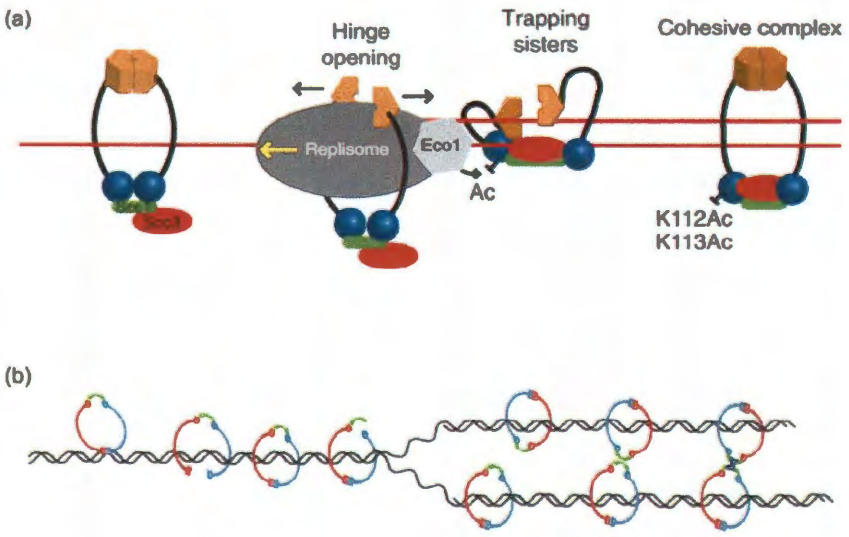


Figure 1.4 Models for the establishment of cohesion between sister chromatids. (a) One-ring model, the hinge domain transiently opens as the replication forks passes through and recloses afterwards (adapted from Kurze *et al.*, 2010). (b) Two-ring model, cohesin rings slide onto one of the chromatids when they pass the replication fork before they are connected by SA1/2 (adapted from Zhang *et al.*, 2009).

The loading of the cohesin complex occurs before the DNA replication. Therefore, question on how the replication fork passes the cohesin complex needs to be answered. An explanation was given for the one-ring embrace model based upon the structural information of the Smc1/Smc3 hinge domain, which contains a positively charged channel with DNA-binding potential (Kurze *et al.*, 2010). When DNA is replicated, the replisome passes the cohesin ring, the Smc1/Smc3 hinge transiently opens and the sister chromatids are co-entrapped by the cohesin ring (Figure 1.4a; Gruber *et al.*, 2006; Kurze *et al.*, 2010). With the opening of the hinge domain, the coiled coil arms bend and the hinge domain transiently associates with the Smc3 NBD, which triggers the acetylation of Smc3. Once Smc3 is acetylated, the hinge domain reassociates with each other, leading to the re-closing of the cohesin ring. It is likely that the acetylation cause the conformation change of Smc3 NBD thus to abrogate the interaction with the hinge domain or to 'straighten' the bent coiled coil arm, leading to the re-closing of the cohesin ring. Furthermore, because the two consecutive acetylated lysine sites are adjacent to the ATP binding pocket, the conformational change may also involve an ATP hydrolysis cycle. However, it is unclear if the coiled-coil arms are flexible enough for the Smc3 hinge domain to 'fold' into the NBD, which requires the 'upper' half of the arm (i.e., the arm part flanking the hinge domain) to fold for almost 180 degrees. Even the Smc3 hinge domain does interact with its NBD, it is unknown whether and how the interaction will trigger the acetylation of the NBD.

In the handcuff model, each ring passes the replication fork by the opening of the Smc3 head and the Rad21 N-terminus and "slide" onto one of the sister chromatids (Figure 1.4b; Yeh et al, 2008). The cohesin rings on the sister chromatids are then paired

and connected by SA1/2 with the possible assistance of other cohesin associated components (Zhang *et al.*, 2008; Zhang and Pati, 2009). Compared to the one-ring model, handcuff model better explains how the cohesin complex passes the replication fork. However, it is unclear how cohesin rings are paired with each other. For the two cohesin rings to pair, first they need to locate at the same locus on each sister chromatid to have a close proximity for interaction. The cohesin complexes are reported to locate at certain loci on sister chromatids (Yeh *et al.*, 2007), which supports the above hypothesis. Second, the two cohesin complexes are required to associate with each other. It has been reported that two Rad21 molecules interact with each other in an anti-parallel manner with the existence of SA1 or SA2 (Zhang *et al.*, 2008). Therefore, it is likely that SA1/2 is the connector for the two cohesin ring. However, it is unclear how the 'connection' takes place. It is not known whether it is a direct interaction or through recruiting other cohesin associated protein. Furthermore, each cohesin carries one molecule of SA1 or SA2, while the handcuff pair only carries one instead of two molecules of SA1/2. Therefore, one of the SA1/SA2 molecules needs to be dissociated from the cohesin complex in the pairing process. However, it is unknown how this dissociation takes place.

1.2.5 The dissociation of the cohesin complex

The dissociation of the cohesin complex from the chromatids differs in yeast and higher eukaryotes. In yeast, the dissociation of all the cohesin complexes takes place at the metaphase to anaphase transition through the anaphase-promoting complex or cyclosome (APC/C)-Separase pathway (Figure 1.5), in which Rad21 is cleaved by an activated endopeptidase (Uhlmann *et al.*, 1999; Uhlmann *et al.*, 2000). Separase is a large

protein (the human Separase is over 200kD). Separase protein contains a supercoiled 26 Armadillo (ARM) repeats in the N-terminus, followed by a unstructured region and two caspase folds, of which only the second one seems to be active through detailed bioinformatic analysis (viadiu *et al.*, 2005). Before the transition from metaphase to anaphase, Separase is bound to its inhibitor, securin (Zou *et al.*, 1999) and remains inactive. The binding is likely extended from the ARM repeats to the caspase fold (Viadiu *et al.*, 2005). At the onset of anaphase, an ubiquitin protein ligase, APC/C, is phosphorylated by a kinase called Cdk1 (Kraft *et al.*, 2003). APC/C is then able to bind to one of its activators, Cdc20, which recruits one of its substrates, securin (reviewed by Peters, 2002). Securin is then degraded and dissociates from Separase. Upon the dissociation of securin, Separase is activated by self-cleavage. Rad21 is then cleaved by the activated Separase at R^{180} and R^{268} (Uhlmann *et al.*, 1999), which leads to the open of the cohesin ring and the dissociation of all the cohesin complexes.

In higher eukaryotes, the dissociation of cohesin complexes takes place in two steps (Figure 1.5; Waizenegger *et al.*, 2000). In prophase, the bulk of the cohesin complexes along the sister chromatid arms are removed through the phosphorylation of SA2 by a polo-like kinase (Plk1) (Sumara *et al.*, 2000; Sumara *et al.*, 2002; Gimenez-Abian *et al.*, 2004). This is called the prophase pathway, which is independent of the APC/C mediated Separase activation. An orthologs of the *D. melanogaster* wings apart-like protein (Wapl) also plays an important role in the removal of arm cohesion possibly by unlocking cohesin from stably binding to chromatin (Kueng *et al.*, 2006). The residual cohesin complex on the arms along with the centromere cohesin complexes are protected by shugoshin (Sgo1) by preventing SA2 from being phosphorylated by Plk1

(McGuinness *et al.*, 2005; Kitajima *et al.*, 2006). Aurora B also regulates the prophase pathway, but its exact role is not known. It may recruit Sgo1 to the centromeric region to protect SA2 during prophase. It is not clear how the prophase pathway works. Since most of the cohesin complexes (not just 'cohesion') are removed from the sister chromatids, it is likely that the process involves the opening of the cohesin ring and the 'escape' of the sister chromatids in both one-ring and handcuff models. Phosphorylation of SA2 might cause conformational change of SA2 which facilitates the opening of the ring. The centromeric cohesin complexes and the rest of the arm cohesin complexes are removed in the APC/C-Separase pathway at the transition from metaphase to anaphase (Hauf *et al.*, 2001; Haering *et al.*, 2003).

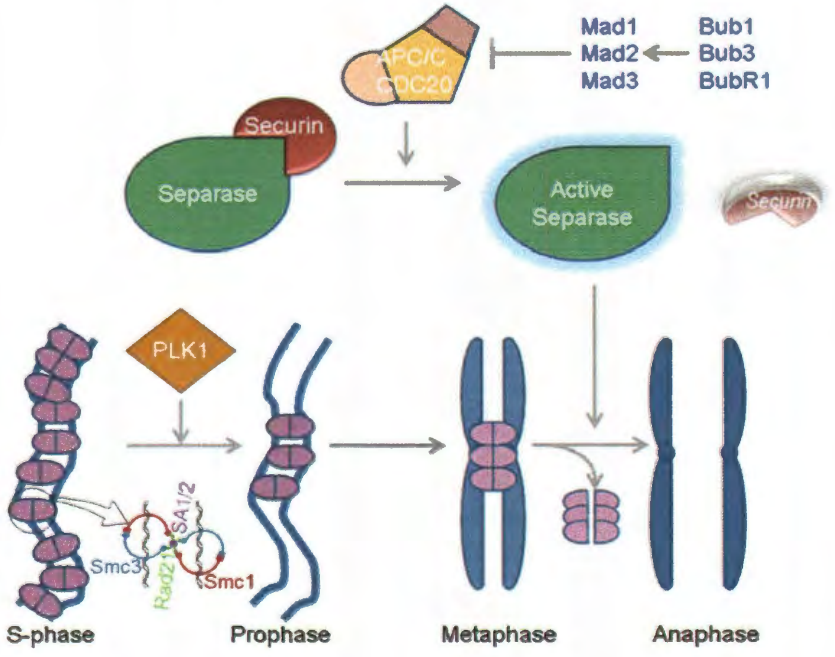


Figure 1.5 APC/C-Separase pathway and prophase pathway (from Panigrahi *et al.*, 2009).

The physiological significance of the prophase pathway is not clear. The APC/C – Separase pathway alone is sufficient for cleaving all the cellular Rad21 (Kueng *et al.*, 2006). One possibility for the necessity of the prophase pathway may be facilitating the recycling of the cohesin complex. It was found that the cohesin complexes dissociated in the prophase pathway are not cleaved by Separase during mitosis and a large portion of them re-associate with the cohesin complex in the telophase (Sun *et al.*, 2009; Waizenegger, *et al.*, 2000; Peters *et al.*, 2008). The purpose of the association might be related to the transcriptional regulation function of the cohesin complex during interphase. For example, the cohesin complex is found to bind to the CCTC-binding factor CTCF, which is a transcription regulator (Rubio *et al.*, 2011). If all the cohesin complex are cleaved by Separase, the Rad21 will need to be re-synthesized, which might lead to the delay of the transcriptional regulation and cause serious consequence for cells.

1.3 Cohesin and Diseases

The accurate duplication and segregation are essential for maintaining the genome integrity for eukaryotes. Deficiency in any of these processes may potentially cause cancer, birth defects and other genetic diseases. Cohesin complex plays an important role in regulating the fidelity of the sister chromatid segregation through providing the cohesion between sister chromatids and being dissociated from the chromatin in a controlled manner.

Previous studies have shown that the deficiency of cohesin and cohesin-related genes are evident in many human cancer and other diseases. Smc3 is overexpressed in

human colon cancer cell lines (Ghiselli *et al.*, 2000). Overexpression of Separase, which cleaves Rad21 at the onset of metaphase, in mammary epithelial cells is sufficient to induce tumorigenesis in a p53 mutant background in mice (Zhang *et al.*, 2008). The overexpression of Separase might cause the premature cleavage of Rad21, and thus lead to aneuploidy. The mutations in Scc2, Smc1a and Smc3 can lead to Cornelia de Lange syndrome which causes developmental disorders with intellectual disability, skeletal abnormality and other developmental deficiencies (Pie *et al.*, 2010).

Studying the human cohesin complex at the molecular basis and the role it plays in genomic instability will provide insights into our understanding of the mechanism of missegregation and contribute to the development of therapeutic methods for cancers and other diseases.

Chapter II. Cloning, expression, purification and structural characterization of SA2 by analytical ultracentrifugation

A robust expression system is essential for preparing the protein of interest for structural and functional studies. The human cohesin complex components are likely to undergo elaborate folding pathways with post-translational modifications. Therefore, the baculovirus expression system, as a eukaryotic expression system, in which insect cells are infected by baculoviruses overexpressing the genes of interest, may benefit the expression of the human cohesin complex subunits. SA1/2 is the fourth core subunit of the cohesion complex. However, other than its supportive role in chromatin cohesion, little is known about its function. Additionally, no structural information on SA proteins is available thus far. In humans, SA2 is more abundant than SA1 (Losada *et al.*, 2000). Therefore, this study mainly focuses on the SA2 protein.

2.1 Cloning of SA2 into insect cells

Insect cells allow many sophisticated post-transcriptional modifications similar to those observed in higher eukaryotes. The relative simplicity and lower cost of the insect cell-baculovirus system make it advantageous over mammalian cell expression system. Thus, insect cells were chosen to express the human cohesin complex subunits. A schematic illustration of how insect cells can be used to express a protein of choice is shown in Figure 2.1. The gene of interest is first cloned into a donor plasmid pFastBac^{TM1} (Invitrogen) carrying a mini-Tn7 element. The recombinant plasmid is then transformed into competent DH10Bac *Escherichia coli* (*E. coli*) cells, which contain a bacmid with a

Tn7 target site and a helper plasmid encoding a transposon. After the transformation, the gene of interest is transposed to the Tn7 target site, leading to the disruption of a *LacZα* gene. Due to this disruption, colonies containing bacmids with the gene of interest (positive) have a white appearance in the presence of Isopropyl β-D-1-thiogalactopyranoside (IPTG) on an X-gal plate as opposed to the blue appearance of colonies containing a functional *lacZα* gene (negative). Recombinant bacmids are then purified and transfected into *Spodoptera frugiperda* cell line 21 (Sf21) insect cells in order to generate the recombinant viruses overexpressing the gene of interest. Finally, the viruses are amplified to reach a higher titer for optimal protein expression.

To clone SA2, the entire cassette containing the SA2 gene with an HA tag at the N-terminus was digested out from 5μg of the pCruz-SA2 construct (pDP871, courtesy of Dr. Debananda Pati) with endonucleases KpnI (NEB) and SacI (NEB) in NEB buffer 1 at 37°C overnight. 3μg of the pFastBac^{TM1} vector (Invitrogen) was digested in the same manner. Both of the digestion products were purified using a QIAquick Gel Extraction kit (Qiagen). The digested SA2 gene product was ligated into the pFastBac vector at 5:1 ratio with T4 ligase (Roche) at room temperature for 30min. The ligation product was transformed into DH5α competent cells (non-commercial) by a 45s heat shock at 42°C before the cells were spread onto agar plates (LB + Amp$^{50μg/ml}$) and incubated at 37 °C overnight. Colonies were inoculated in a 3ml culture (LB + Amp$^{50μg/ml}$) and the liquid cultures were shaken at 250 rpm in a 37°C incubator overnight. Plasmids were purified using a QIAquick Miniprep Kit (QIAGEN). Positive clones (clones containing SA2 gene) were identified in a 1% agarose gel following a double digestion with KpnI/SacI. The

positive clones were sequenced to verify the proper insertion of the SA2 sequence (sequencing was performed by Seqwright, Inc.).

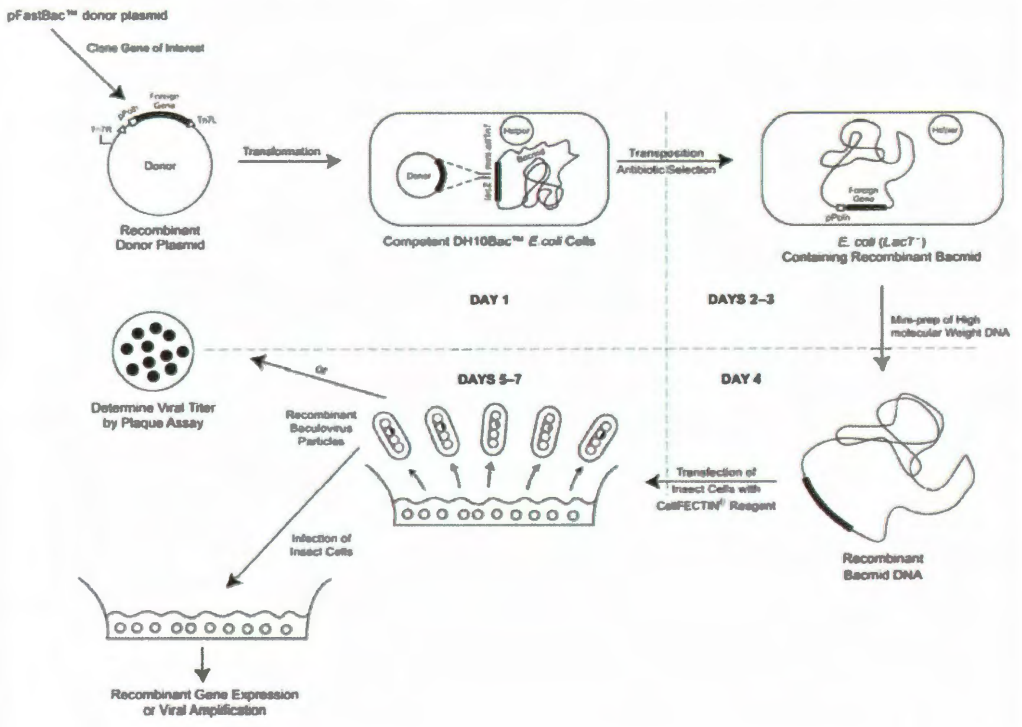


Figure 2.1 Schematic flowcharts showing the generation of baculovirus expressing the genes of interest (Bac-to-Bac® expression system manual, Version D, Invitrogen).

The recombinant plasmid containing the coding sequence of SA2 was transformed into competent DH10Bac *E. coli* cells (Invitrogen). After being heat shocked, DH10Bac cells were shaken at 37°C for 4h and then spread onto X-gal agar plates (LB + Kan50$^{\mu g/ml}$ + Tet$^{10\mu g/ml}$ + gentamicin$^{7\mu g/ml}$ + IPTG$^{40\mu\ g/ml}$ + X-gal$^{40\mu g/ml}$). After 48h of incubation at 37°C, several white colonies were selected and re-streaked onto

fresh X-gal plates followed by incubation at 37°C for 48h. Once confirmed by restreaking, white colonies were selected and inoculated into a 5ml culture (LB + Kan50$^{\mu g/ml}$ + Tet$^{10\mu g/ml}$ + gentamicin$^{7\mu g/ml}$), which was then vigorously shaken at 37°C for 16h. Bacmids were extracted using a modified alkaline lysis miniprep protocol as described in the Bac-to-Bac® baculovirus expression system manual (Version D, Invitrogen). Polymerase Chain Reaction (PCR) with M13 forward and reverse primers as described in the Bac-to-Bac® baculovirus expression system manual was used to verify that the bacmids contained the gene of interest. The size of the extension product was found to be ~6 kb, consistent with the expected size of the SA2 gene (3.7kb) plus 2.3kb (including the gentamycin resistant gene (*Gent*^{R}), which is reversely orientated in the pFastBac vector, and some extra sequence from the bacmid DNA).

The concentrations of the purified recombinant bacmids were measured using a DyNA quant 200 fluorometer (Hoefer). The concentration of a purified bacmid is 100ng/μl. To ensure optimal transfection efficiency, approximately 9 x 10^5 cells Sf21 cells were seeded per well in a 6-well tissue culture plate (Corning). The plate was then incubated at 27 °C for 40min to allow the formation of a monolayer of cells attached to the bottom of the wells. For each of the three positive recombinant plasmids selected for transfection, 1μg of DNA and 6μl of Cellfectin® II Reagent (Invitrogen) were individually diluted in 100μl of un-supplemented Grace's insect cell medium (Sigma-Aldrich) before they were gently mixed together. The DNA-Cellfectin mixture was then incubated at room temperature for 30min. The purpose of this process is to form a DNA-lipid complex for optimal transfection efficiency. The seeded cells were then taken out of the incubator and the attachment of the cells to the bottom of the wells was confirmed

using a light microscope (VWR). The medium in each well was discarded and each well was washed with 2ml of un-supplemented Grace's medium. For each transfection, the 200μl of DNA-lipid complex was diluted in 0.8ml of un-supplemented Grace's medium before being added to each well. After incubation for 5h at 27 °C, the un-supplemented medium in each well was replaced with 2ml of Grace's medium supplemented with 10% Fetal Bovine Serum (FBS; Invitrogen) before incubation at 27°C for 5~7 days. During the course of the incubation, the cells were monitored daily for signs of infection, including swelling and detaching from the bottom of the well. To harvest the P1 virus, cells from each well were spun down at 800g for 10min and the supernatants were saved as the P1 viral stock. The cell debris remaining in the pellets was examined by SDS-PAGE to determine the expression level of the protein. Cell pellets were resuspended in a hypotonic lysis buffer (50 mM Tris-HCl, pH=7.5, 50mM NaCl, 1% Triton X-100 and 1mM PMSF) and incubated on ice for 30min. The lysates were then centrifuged at 18,000g for 15min. The supernatant was electrophoresed on an 8% SDS-PAGE gel and subsequently stained in a Coomassie blue solution (Gelcode). A strong band for the HA tagged SA2 protein was visible in a Coomassie stained gel, indicating HA-SA2 is properly expressed in Sf21 cells and the expression level could be sufficient for structural and functional studies.

For virus amplification, the P1 viral stock was used to infect Sf21 cells at a multiplicity of infection (MOI) of 0.05 to generate a P2 viral stock and subsequent P3 viral stock. The titer of each viral stock was determined by plaque assays as described in the Bac-to-Bac® Baculovirus Expression System Manual (Version D, Invitrogen). The

titer of the P3 viral stock was typically above 10^8 pfu/ml. The virus stocks were protected from light exposure using storage bottles wrapped in foil and stored in a 4°C refrigerator.

2.2 Purification of HA-SA2 and His-SA2

To purify HA-SA2 from sf21 cells, ~2 x 10^9 (or 2 liters) cells were infected with the recombinant baculovirus at an MOI of 5 and collected 48h post-infection. The cell pellets were then washed with phosphate-buffered saline (PBS; 137mM NaCl, 2.7mM KCl, 8.1mM Na_2HPO_4 and 1.5mM KH_2PO_4, pH=7.5). Cell pellets were resuspended in 35ml of Tris buffered saline (TBS, 50mM Tris-HCl, 150mM NaCl, pH=7.5) and lysed by sonication using an Ultrasonics Heat Systems sonicator model W-375 at power level 7 (60% duty cycle) for 3 x 2.5min. The sonication process was performed on ice to prevent overheating which might lead to the degradation and denaturation of cellular proteins. The cell lysate was then centrifuged at 35,000g for 40min at 4 °C using a Beckman centrifuge and the supernatant, which contained >90% of the total HA-SA2, was collected for further purification.

As the first purification step, the ammonium sulfate precipitation method was used instead of the more expensive and less efficient immunoaffinity purification using anti-HA monoclonal antibody (mAb) conjugated beads. A small scale ammonium sulfate gradient precipitation test was performed to demonstrate that SA2 precipitates at 20% - 30% (w/v) ammonium sulfate concentration. The total volume of the supernatant was then measured and ammonium sulfate was added to a final concentration of 20% (w/v). The mixture was gently rotated at 4°C for an hour and the supernatant was obtained after a 35,000g x 30min centrifugation. The volume of the supernatant was then measured and

ammonium sulfate was added to a final concentration of 30% (w/v), after which the mixture was gently rotated for 1h at 4°C. The final supernatant was discarded and the pellets were resuspended in 30ml of FPLC buffer A (50mM Tris-HCl, pH=7.5, 1 mM EDTA, 2mM βME, and 1 mM NaN_3). The ammonium sulfate precipitation fractions were analyzed by an 8% SDS-PAGE gel. In addition to a ~140kD band expected for HA-SA2, a strong ~120kD product was also observed, which might be a degradation product of the HA-SA2 protein (Figure 2.2a). To remove any protein aggregates and cellular debris that may cause potential harm to chromatography columns, the sample was spun at 35,000g for 30min and the supernatant was used for subsequent chromatography purification.

To further purify HA-SA2, an ion-exchange column was used. HA-SA2, with a theoretical isoelectric point (pI) = 4.27, was loaded onto a 5ml Hitrap Q column (GE Healthcare) which is an anion exchange column that selectively binds negatively charged proteins. Proteins were eluted at an increasing gradient of NaCl from 50mM to 1M at pH 7.5. HA-SA2 was eluted at about 400mM NaCl. Fractions from the Q column were analyzed by an 8% SDS-PAGE gel and the ~120kD product was present in almost every fraction that contained the full length SA2 (Figure 2.2 b).

Samples containing the HA-SA2 protein were concentrated with a centricon tube (Millipore) to 2ml before being loaded onto a 120ml Superdex 200 column (GE Healthcare), which is a gel filtration column that separates proteins by molecular weight (MW) and shape. HA-SA2 was eluted with a gel filtration buffer (50mM Tris-HCl, 200mM NaCl, 10% Glycerol, 1mM EDTA, 2mM β-mE, 1mM NaN_3) at an MW position slightly larger than the 150kD MW standard (Figure 2.2c). Fractions from the gel

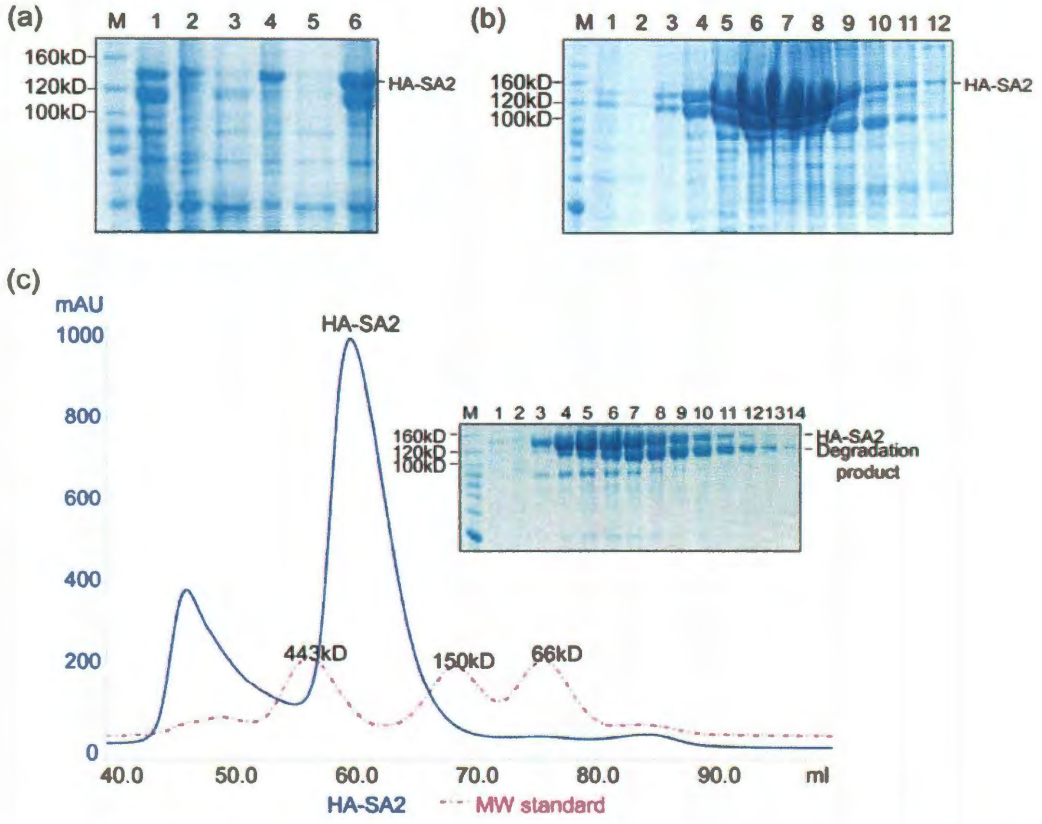


Figure 2.2 Purification of HA-SA2 from insect cells. (a) Fractions from ammonium sulfate precipitation. M: molecular weight marker, lane 1, supernatants of the cell lysates; lane 2, pellets of the cell lysates; lane 3, pellet from precipitated by 20% ammonium sulfate dissolved in FPLC buffer A; lane 4, supernatant from 20% ammonium sulfate precipitation; lane 5, supernatant of from 30% ammonium sulfate precipitation; lane 6, pellet from 30% ammonium sulfate dissolved in FPLC buffer A. (b) Fractions from the Q column. Lane 1, flowthrough; lanes 2-12, fractions eluted from the Q column. Fractions from lanes 5-8 were collected for gel filtration. (c) The gel filtration chromatograph and fractions from the gel filtration column. Lane 1, samples from the aggregation peak; lanes 2-14, fractions from the major sample peak.

filtration chromatogram were analyzed using an 8% SDS-PAGE gel which showed that SA2, the ~140kD protein, had been co-eluted with the ~120kD molecule, although the ~120kD protein migrated a little slower than the ~140kD counterpart. The small difference between the MW of the two proteins made it difficult to separate them.

The failure to purify the full length SA2 protein to homogeneity could have been due to the instability of the HA-SA2 protein or alternatively the lengthy ammonium sulfate precipitation that might result in protein degradation (Figure 2.2a, Lane 4 vs. Lane 6). To ease the initial purification process, a six-histidine (6xHis) tag was introduced to the N-terminus of SA2 by re-cloning using a 5' primer containing six histidine codons at the very 5' end of the SA2 cDNA sequence. To insert the gene of SA2 with a 6xHis tag at the N-terminus into the pFastBac^{TM1} vector, a PCR reaction was performed to amplify the SA2 gene. The 50µl PCR mixture was composed of 0.5µl of template at 10ng/µl (HA-SA2), 0.5µl of pfu Turbo polymerase (Stratagene), 5µl of 10 x pfu Turbo reaction buffer (Stratagene), 1.5µl of dNTPs at 100mM (Promega), 5µl of forward primer at 10µM, 5µl of reverse primer at 10µM, and 32.5µl of Milli-Q (MQ) water. The PCR reaction was performed in a Mastercycler® Pro S (VWR) PCR machine starting with a 5min denaturation step at 94 °C, followed by 30 cycles of denaturation (94°C for 45s), annealing (50 °C for 45s), and extension (72°C for 8min). This was followed by a final extension step at 72°C for 10min to ensure complete extension of duplicate DNA strands. The PCR product was analyzed by electrophoresis in a 0.8% agarose gel. A DNA fragment of approximately 3700 bp was obtained. The PCR product was purified with a QIAquick PCR Purification Kit (Qiagen). Both the PCR product and the pFastBac^{TM1} vector were digested with restriction enzymes BamHI (NEB) and XhoI (NEB) in an NEB

buffer at 37 °C overnight. The PCR products were then ligated into the pFastBac^{TM1} vector. The subsequent cloning procedures were the same as the one described in section 2.1 for producing higher titer baculoviruses overexpressing the 6 x His tagged SA2.

To purify the 6xHis tagged SA2, ~2 x 10^8 Sf21 cells were transfected with baculoviruses overexpressing His-SA2 and harvested 48h post-infection. Cell pellets were resuspended in 10ml of Ni-NTA buffer A (50mM Tris-HCl, pH 7.5, 300mM NaCl, 20mM imidazole, 10% glycerol, 0.1mM PMSF) and lysed by sonication. Supernatants from the lysates were collected by 35,000g x 40min centrifugation and loaded onto 0.2ml of pre-equilibrated Ni-NTA resin (Qiagen). The supernatant was then constantly mixed with the resin at 4°C for 1h. The resin was collected by 1,000g x 3min centrifugation and the supernatant (flow through) was removed. The resin was then washed with 5ml of Ni-NTA Buffer A. The proteins bound to the Ni-NTA agarose were eluted by Ni-NTA buffer B (50mM Tris-HCl, pH 7.5, 300mM NaCl, 250mM imidazole, 10% glycerol, 0.1mM PMSF). The eluted fractions were analyzed by an 8% SDS-PAGE gel (Figure 2.3a, Lanes 1-5). An ~120kD degradation product was also present in the Ni-NTA purified fractions. Compared to the initial purification of HA-SA2, the amount of the degradation product present decreased significantly.

The fractions containing the His-SA2 protein were collected and spun down at 18,000g for 10min for a complete removal of aggregates and cellular debris. To further purify the protein, an anion-exchange column (Figure 2.3a, Lanes 6-12) and a gel filtration column were used as described in the purification of HA-SA2. The sizes of the columns used were different due to the small scale of the purification performed. A 1ml Q column (GE Healthcare) was used instead of a 5ml Q column and a 24ml Superose 6

column was used instead of a 120ml superdex 200 column. The fractions from the gel filtration column are shown in Figure 2.3b. The ~120kD product was again present in every fraction that contained the full length His-SA2.

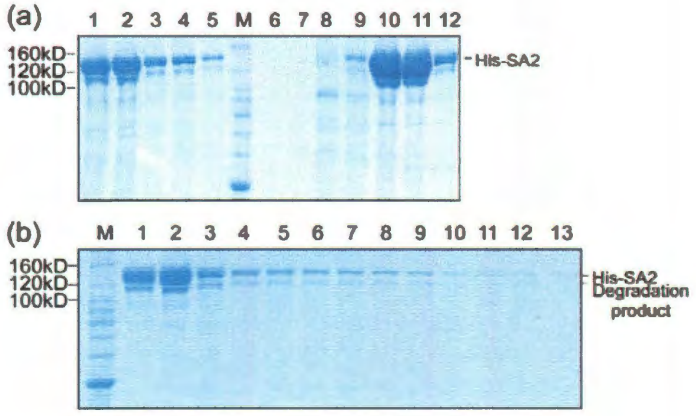


Figure 2.3 Purification of His-SA2 from Sf21 cells. (a) Fractions from the Ni-NTA column and the Q column. Lanes 1-5, fractions eluted from the Ni-NTA column; lane 6, flow through from the Q column; lanes 7-12, fractions eluted from the Q column. Lanes 10-11 were collected for gel filtration. (b) The fractions from the gel filtration column (lanes 1-13).

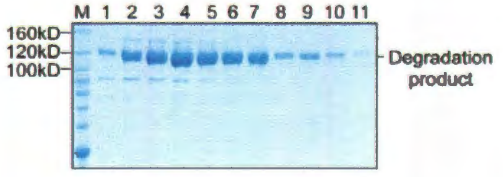


Figure 2.4 HA-SA2 treated at room temperature for 8 days. Lanes 1-11 were the same samples from Lanes 3-13 in Figure 2.3c.

2.3 Expression and purification of two SA2 deletion mutants

Crystallization studies require highly pure and homogenous protein samples. The full length SA2 proteins purified from section 2.2 contained a degradation product, which made the sample impure and unsuitable for crystallization. Hence, it was necessary to make a more stable SA2 construct for crystallization studies.

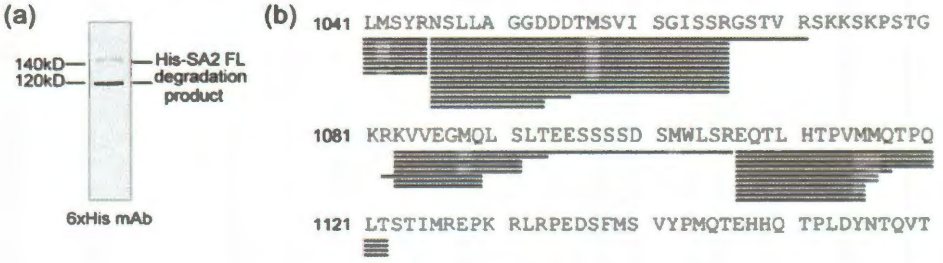


Figure 2.5 Identification of the SA2 degradation product. (a) Western blot of the purified SA2 full length protein. The 6xHis tag at the N-terminal of the full length SA2 was detected by 6xHis mAb. The ~120kDa degradation product also contained the N-terminal tag. (b) In-gel digestion and peptide identification by HPLC/MS. T^{1122} is the last amino acid identified by MS. The peptide coverage at the N-terminal region is not shown, as the 120kDa product contains the 6xHis tag and thus should have all the N-terminal sequence.

We found that the ~120kD degradation product of the full length SA2 was stable. When the fractions from Figure 2.2c (Lanes 3-13) were placed at room temperature for 8 days and analyzed by an 8% SDS-PAGE gel (Figure 2.4), the full length SA2 was almost completely degraded to the ~120kD product. The high stability of the ~120kD product thus made a potentially ideal candidate for crystallization studies. Western blot using the

anti 6xHis monoclonal antibody (mAb; BD biosciences) showed that both the full length SA2 and the degradation product contained the N-terminal 6xHis tag (Figure 2.5a). Therefore, the degradation of SA2 must have occurred near the C-terminus. Mass Spectrometry (MS) was used to identify the cleavage site. The ~120kD product band was isolated from a Coomassie stained gel, washed in MQ water for 3 x 5min and sent to the Tufts Core Facility for MS analysis. Peptides from an in-gel digestion by trypsin and chymotrypsin were also analyzed by High performance liquid chromatography (HPLC)-MS and the results showed that amino acids 1~1122 were present (Figure 2.5b), suggesting that the cleavage occurred right after T^{1122}, leading to the design of an SA2 deletion mutant, SA2 (1-1122aa). However, a secondary structure prediction using PredictProtein (Rost *et al.*, 2004) showed that there was an extended unstructured region flanking the site T^{1122} (Figure 2.6). Considering that the unstructured region may be flexible, thus to cause problems in crystallization, we made two SA2 deletion mutants: one is SA2 (1-1122aa), the same as the ~ 120kD degradation product and the other is SA2 (1-1051aa), a slightly shorter construct in which the unstructured region at the C-terminus was genetically removed.

Both of the deletion mutants SA2 (1-1051aa) and SA2 (1-1122aa) were individually expressed in ~2 x 10^9 Sf21 cells and purified using Ni-NTA resin, as well as anion exchange and gel filtration columns similar to the purification of the full length His-SA2 (Figure 2.7 & Figure 2.8). A 5ml Q column and a Superdex 200 column were used in the purification. Both proteins had an apparent MW that was slightly greater than 150kD (Figure 2.7c & Figure 2.8c). The difference between the theoretical MW (~120kD) and the apparent MW from the gel filtration column can either be the result of

dimerization, yielding a theoretical MW to ~240kD or a non-spherical shape for both mutants. This size issue of SA2 will be addressed in more details in section 4.4. The purified SA2 deletion mutants were at least 95% pure as determined by SDS-PAGE (Figure 2.7c & Figure 2.8c). Both deletion mutants were then concentrated to 5mg/ml for crystallization studies. The concentration was determined using both Bradford assay and A_{280} measurements.

```
                  970         980         990        1000        1010
         ....|....| ....|....| ....|....| ....|....| ....|....|
SA2      LDQLKTREAI AMLHKDGIEF AFKEPNPQGE SHPPLNLAFL DILSEFSSKL
PROF_Sec ----HHHHHH HHHHHHHHHE EEE------- ------HHHH HHHHHHHHHH
REL_Sec  5630014678 8876631111 1203466666 6665321468 8888877774

                 1020        1030        1040        1050        1060
         ....|....| ....|....| ....|....| ....|....| ....|....|
SA2      LRQDKRTVYV YLEKFMTFQM SLRREDVWLP LMSYRNSLLA GGDDDTMSVI
PROF_Sec HH  HHHHHH HHHHHHH              HHHHHHHHH
REL_Sec  2003578888 7876410235 5455333320 3567664300 3334565434

                 1070        1080        1090        1100        1110
         ....|....| ....|....| ....|....| ....|....| ....|....|
SA2      SGISSRGSTV RSKKSKPSTG KRKVVEGMQL SLTEESSSSD SMWLSREQTL
PROF_Sec
REL_Sec  5545455434 5666777533 3455666555 6666555665 5444322334

                 1120        1130        1140
         ....|....| ....|....| ....|....|
SA2      HTPVMMQTPQ LTSTIMREPK RLRPEDSFMS
PROF_Sec
REL_Sec  5666665666 5543333544 4566444456
```

Figure 2.6 Secondary structure prediction for SA2. SA2 amino acid sequence is colored based on residue types (e.g. blue for positively charged, red for negatively charged, green for hydrophobic, silver for polar, etc). PROF_sec predicts the secondary structure (H = Helix, E = strand). Rel_Sec shows the reliability index of the PROF_sec prediction (0=low, 9=high).

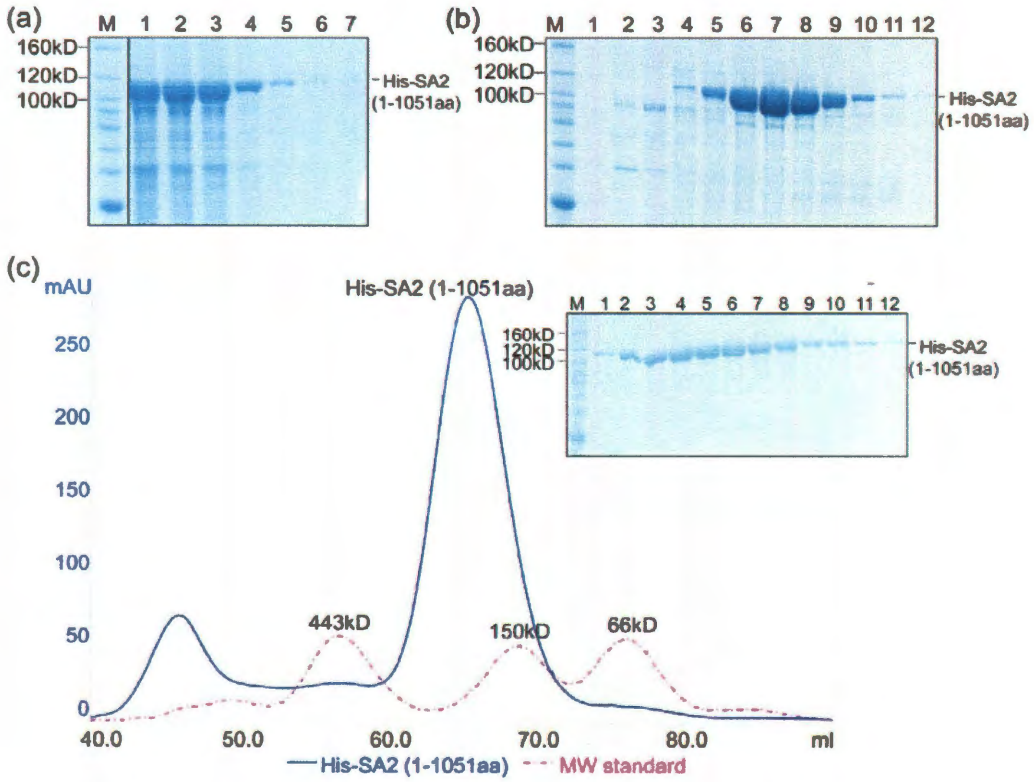


Figure 2.7 Purification of SA2 (1-1051aa) from Sf21 cells. (a) Fractions from the Ni-NTA column. (b) Fractions from the Q column. Lane 1, flow through from the Q column; lanes 2-12, fractions eluted from the Q column. Fractions from lanes 6-9 were collected for gel filtration purification. (c) The gel filtration chromatogram and fractions from the gel filtration column. Lanes 1-12, peak fractions from gel filtration.

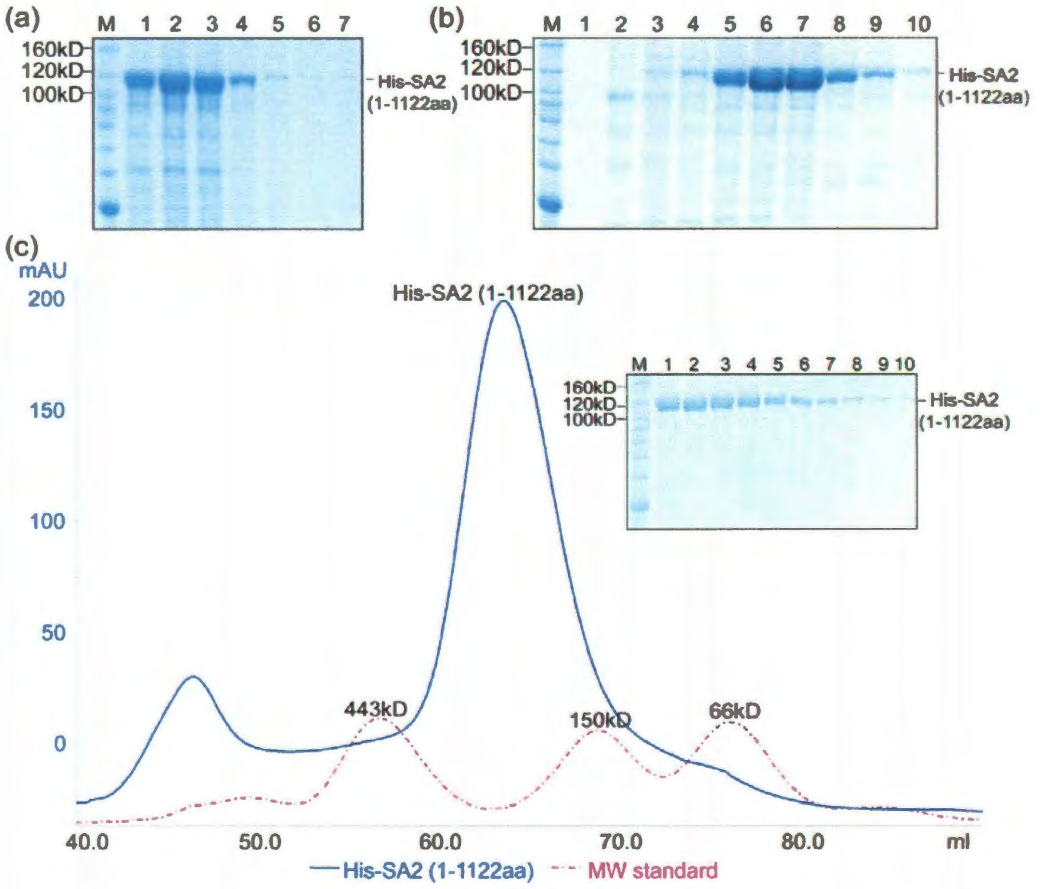


Figure 2.8 Purification of SA2 (1-1122aa) from insect cells. (a) Fractions from the Ni-NTA column. (b) Fractions from the Q column. Lane 1, flow through from the Q column; lanes 2-10, fractions eluted from the Q column. Fractions from lanes 5-8 were collected for gel filtration purification. (c) The gel filtration chromatogram and fractions from the gel filtration column. Lanes 1-10, peak fractions from the gel filtration column.

2.4 Attempts to crystallize the SA2 deletion mutants

Crystallization is a process by which crystals are formed from a macromolecular solution. During this process, homogenous molecules form highly ordered three dimensional arrays. Protein crystallization is usually a slow process with controlled precipitation of proteins in an aqueous solution under the 'assistance' of a precipitant, such as ionic compounds or polyethylene glycol (PEG). The process of protein crystallization is summarized in Figure 2.9. The target protein is usually mixed with a precipitant until it reaches the supersaturated state. In the labile zone, under favorable crystallization condition, the primary nucleation of the protein leads to the formation of a microscopic crystal nucleus. The micro crystals will then slowly grow into full size crystal in the metastable zone. However, if the concentration of the protein is too high, the proteins are likely to precipitate or form amorphous aggregates (Ashrie, 2004).

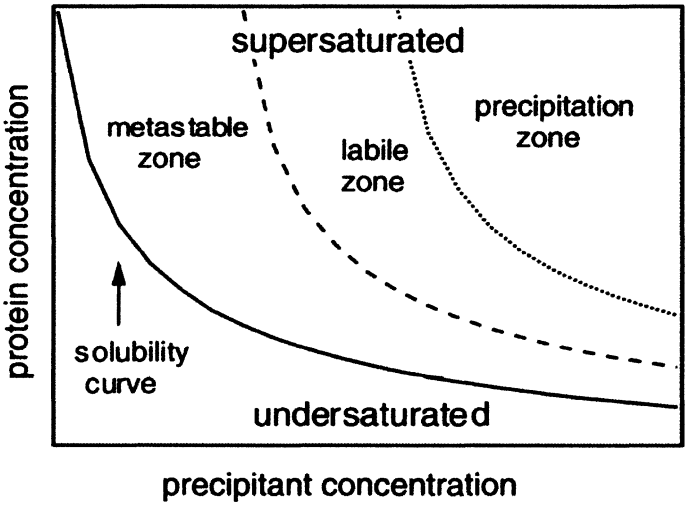


Figure 2.9 A schematic phase diagram depicts the protein solubility as a function of precipitant concentration (from Asherie, 2004).

Protein crystallization is a delicate process where any slight change in the buffer condition, temperature or other external factors may result in amorphous aggregates, poor crystal quality or no crystal at all. In addition to the external factors, the intrinsic properties of the protein sample (e.g., homogeneity, protein stability, flexible region of the protein) have large effects on the formation of crystals.

A large-scale screen is usually the first step in determining the initial crystallization conditions. Commercial screening kits (Qiagen, Hampton Research and Emerald Biosystems) with conditions varying in salt contents, pH, and the types of precipitants were used for the initial crystallization screening. High throughput crystallization screening uses 96-well screening trays (Corning), which are designed to mimic the sitting-drop method. Each well of the 96-well trays holds 100 μl of solution (i.e., mother liquor) and a small dent in the upper edge of the well accommodates a drop of solution composed of an equal amount of the protein and the mother liquor (0.8~1 μl each). A Hydra Plus II robot (Rigaku) was used to set up the trays before they were sealed by a crystal clear tape to prevent evaporation. In the crystallization trays, vapor diffusion is only allowed between the mother liquor and the sitting drop in each well. Around half of the drops in each 96-well tray form precipitants when the protein concentration is about 5mg/ml for both mutants. 28 trays were set up using both SA2 mutants at the concentration of 5mg/ml (14 different kits, each condition was repeated twice). Unfortunately, no 'hits' were observed in the initial screening of either of the SA2 deletion mutants. Considering that our collection of crystallization screening solutions may not cover a very broad range of conditions, SA2 (1-1051aa) was also sent to the Hauptman Woodward Research Institute for more crystallization screening. They

provided 1536 crystallization 'cocktail' conditions using microbatch-under-oil method, in which a very small volume of both sample and cocktail solutions were used (200nl protein + 200 nl cocktail) without the use of mother liquor (Hauptman Woodward Research Institute). However, no crystals were observed in a 6-week long observation.

2.5 Limited digestion of the SA2 protein

The SA2 deletion mutants did not crystallize, perhaps due to the flexibility of the molecules. To identify a rigid core of SA2 for crystallization, the full length SA2 was digested with trypsin and chymotrypsin. A ~70kD band was observed when chymotrypsin was used for limited digestion (Figure 2.10). However, only a very small amount of this protein (< less than 10%) was left in the digested sample, making it difficult to purify this ~70kD product directly from the digestion product.

To prepare the ~70kD digestion product for crystallization, the amino acid composition of the molecule must be determined for cloning and expression. A Western blot using an anti 6xHis antibody showed that the 70kD product had an intact N-terminus. The molecular weight of the ~70kD product was determined accurately by matrix-assisted laser desorption/ionization mass spectroscopy (MALDI-MS). Because the sequence of His-SA2 is known and the digestion product contains the intact N-terminus of His-SA2, the cleavage site was then calculated based upon its accurate MW. The total MW of His tagged SA2 (1~589aa) matches the MW obtained from MALDI-MS. Therefore, the cleavage site was identified as L^{589}. An SA2 mutant containing the first 587 amino acids of SA2 (SA2 (1-587aa)) was designed excluding the two hydrophobic

leucines at the C-terminal that could have negatively affect the solubility of the protein. A baculovirus overexpressing SA2 (1-587aa) was constructed. Unfortunately, SA2 (1-587aa) mostly remained in the insoluble fraction in spite of extensive efforts to solubilize the protein in different solvent conditions (Figure 2.10 b).

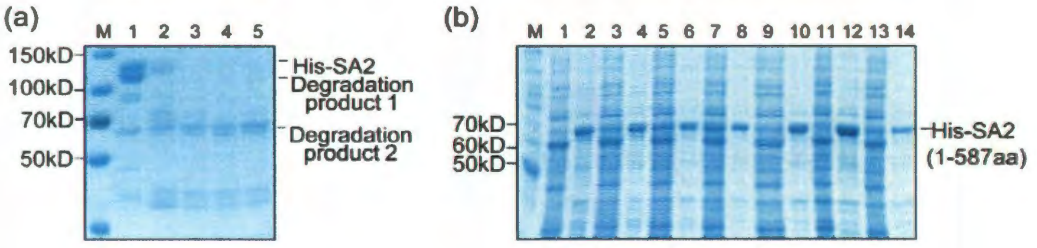


Figure 2.10 Identification and expression of the SA2 rigid core. (a) Chymotrypsin digestion of His-SA2. His-SA2 and chymotrypsin were mixed at the 50:1 mass ratio. Lane 1, SA2 sample before digestion; lanes 2-5: His-SA2 was digested for 1min, 5min, 10min and 20min respectively. Degradation product 2 was relatively stable when treated with chymotrypsin. (b) Purification test for His-SA2 (1-587aa). Sf21 cells expressing His-SA2 (1-587aa) were lysed by sonication in different buffer conditions. The odd number lanes represent supernatants of the cell lysates while the even number lanes represent cell pellets. All lysis buffers contained: Tris-HCl 50mM, pH=7.5. In addition, each individual condition contained: lanes 1-2: 300mM NaCl; lanes 3-4: 200mM $(NH_4)_2SO_4$; lanes 5-6: 500mM NaCl; lanes 7-8: 400mM LiCl; lanes 9-10: 300mM NaCl+1% Glycine; lanes 11-12: 300mM NaCl + 1M glucose; Lanes 13-14: 300mM NaCl+ 60μM Tween20. These additional compounds were previously known to enhance protein solubility (Bondos *et al.*, 2003).

The poor solubility of the SA2 (1-587aa) construct is perhaps due to its inability to form a structured domain by itself. It is also possible that the 588-1051aa region is important for the correct folding of SA2 (1-587aa) during the expression and post-transcriptional modification process. It is also possible that SA2 (1-587aa) requires other peptides to stabilize it after expression. It appeared that the ~70kD protein co-existed with two fragments at ~30-40kD range (Figure 2.10a). It is possible that the SA2 protein was nicked by chymotrypsin at L^{589}, but the SA2 (1-587aa) remained non-covalently associated with one or two smaller peptide species, without which, SA2 (1-587aa) can not stay soluble.

2.6 The cloning, expression, purification, and attempts to crystallize more SA2 deletion mutants

Since our rational design of the SA2 core, SA2 (1-587aa), failed, we decided to employ a systematic approach to search for an SA2 deletion mutant that is soluble and can be used for crystallization.

Baculovirus overexpressing progressive SA2 deletion mutants with 150 amino acids increments/decrements from either the N- or the C-terminus of SA2 (1-1051aa) were generated (Figure 2.11a). In addition to the above deletion mutants, baculovirus overexpressing SA2 (1052-1231aa) was also obtained. To evaluate the expression level and solubility of the various SA2 deletion mutants, the cell pellets separated from the P1 viral stocks were lysed with a hypotonic lysis buffer as mentioned in section 2.1 and the resulting supernatants were analyzed by western blot using an anti 6xHis mAb. The results showed a relatively high expression level of soluble proteins for the deletion

mutants: SA2 (1-150aa), SA2 (1-302aa), SA2 (1-750aa), SA2 (1-903aa), SA2 (451-1051aa), SA2 (581-1051aa), and SA2 (1052-1231aa). These baculoviruses were used for further amplification. Approximately 2 x 10^8 sf21 cells were infected with baculoviruses overexpressing each of the deletion mutants mentioned above at an MOI of 5. Each of the deletion mutants were purified in the order of Ni-NTA column, anion-exchange, and gel filtration columns. The fractions from the gel filtration for each mutant are shown in Figure 2.11b. Among all the deletion mutants tested, only SA2 (1-302aa) and SA2 (451-1051aa) had sufficient yields for crystallization study.

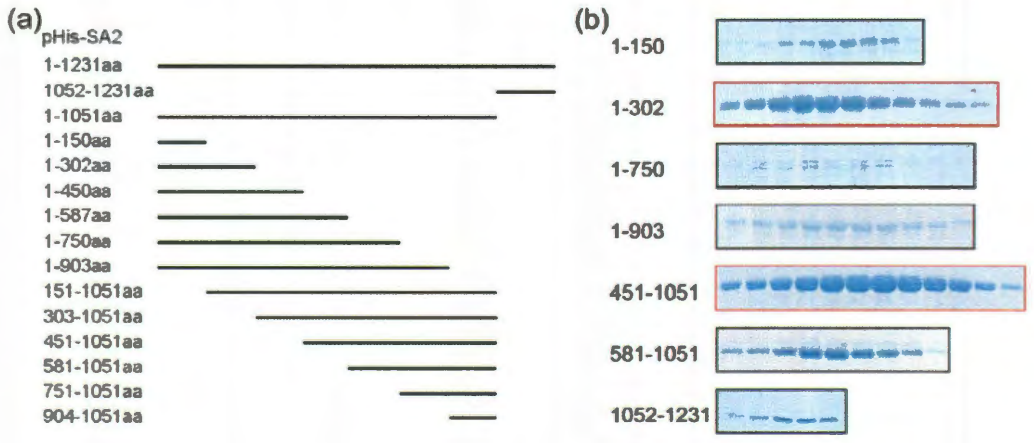


Figure 2.11 Purification test for SA2 deletion mutants. (a) Schematic figure of the SA2 deletion mutants. (b) Fractions from gel filtration for each of the SA2 deletion mutants were analyzed by SDS-PAGE gels. SA2 (1-302aa) and SA2 (451-1051aa) displayed high yield suitable for crystallization studies.

For preparative scale purification, SA2 (1-302aa) and SA2 (451-1051aa) were purified from ~2 x 10^9 cells each as described in section 2.2. The gel filtration chromatograms are shown in Figure 2.12. The apparent MW of SA2 (1-302aa) is ~100kD, indicating the formation of a spherical trimer or a monomer/dimer with non-ideal shapes. The final purified SA2 (1-302aa) sample contained a contaminating ~25kD protein (Figure 2.12a), which might cause some problem in crystallization. In addition, the (1-302aa) protein quickly precipitated at both 4°C and room temperature.

The gel filtration chromatogram showed that SA2 (451-1051aa) was eluted in three forms (Figure 2.12b). By comparing the chromatogram to the MW standard, the three forms are most likely tetramer, dimer and monomer, among which, the dimer was most dominant. Fractions containing the dimer were collected and concentrated to 5 mg/ml for crystallization screenings. Crystallization screenings were both performed in the lab and at the Hauptman Woodward Research Institute. Unfortunately, the protein precipitated heavily, even at 2mg/ml and no crystals were observed.

2.7 The cloning, expression, purification, and attempts to crystallize the SA1 protein

The human SA1 protein shares 70% sequence identity with its ortholog SA2. Considering that SA1 may be a better crystallization target, baculovirus overexpressing the 6xHis tagged SA1 protein was generated and the purification of SA1 was performed as described in 2.2. Resembling SA2, SA1 produced a ~120kD degradation product during the purification process (Figure 2.13 a,b). An in-gel digestion by chymotrypsin followed by an MS analysis showed that the degradation occurred near T^{1123}, which is conserved and equivalent to the previously identified T^{1122} of SA2 (Figure 2.14).

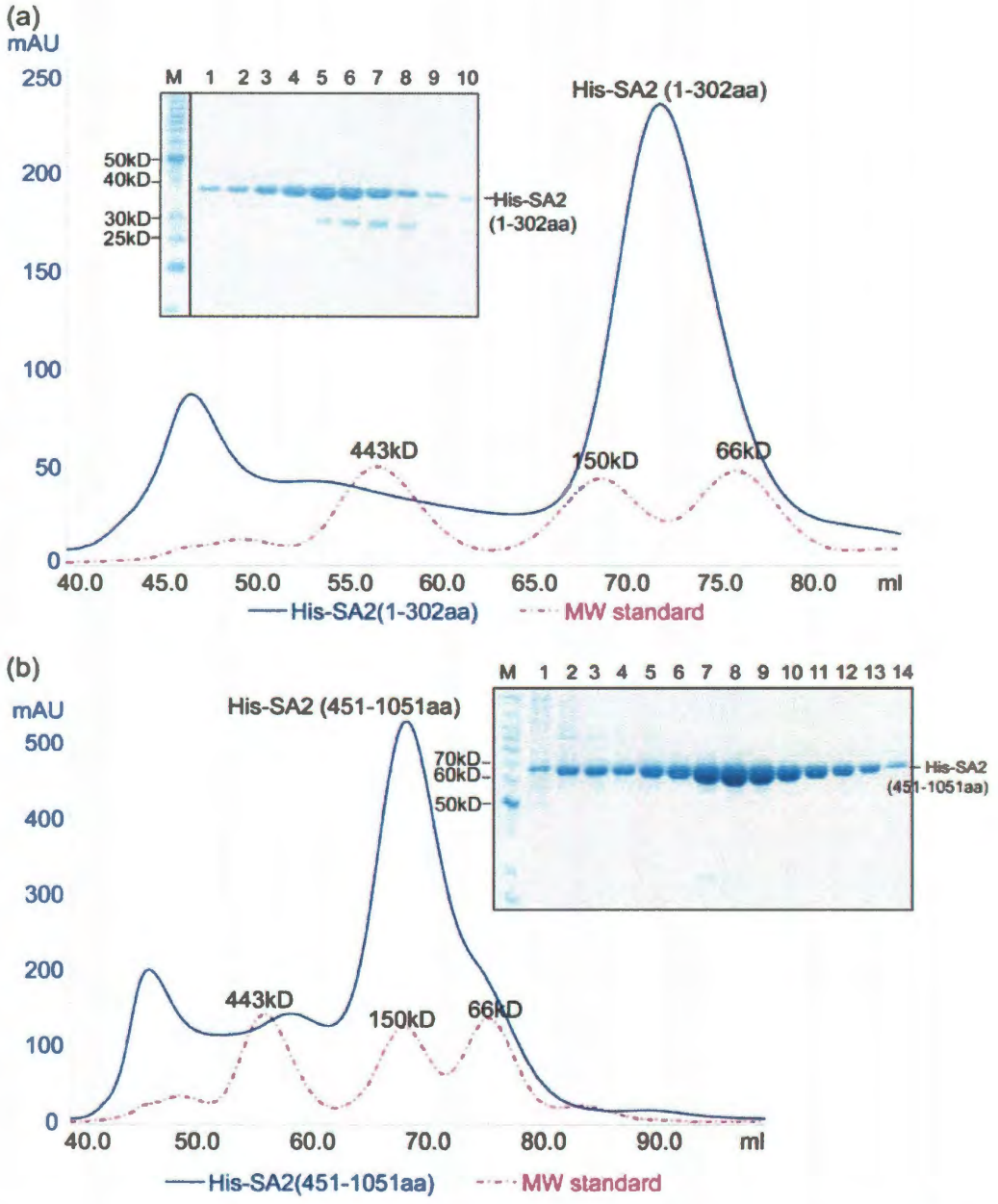


Figure 2.12 Purification of His-SA2 (1-302aa) and His-SA2 (451-1051aa). (a) Gel filtration chromatogram for His-SA2 (1-302aa) and fractions from the gel filtraion (lanes 1-10). (b) Gel filtration chromatogram for His-SA2 (451-1051aa) and fractions from the gel filtraion (lanes 1-14).

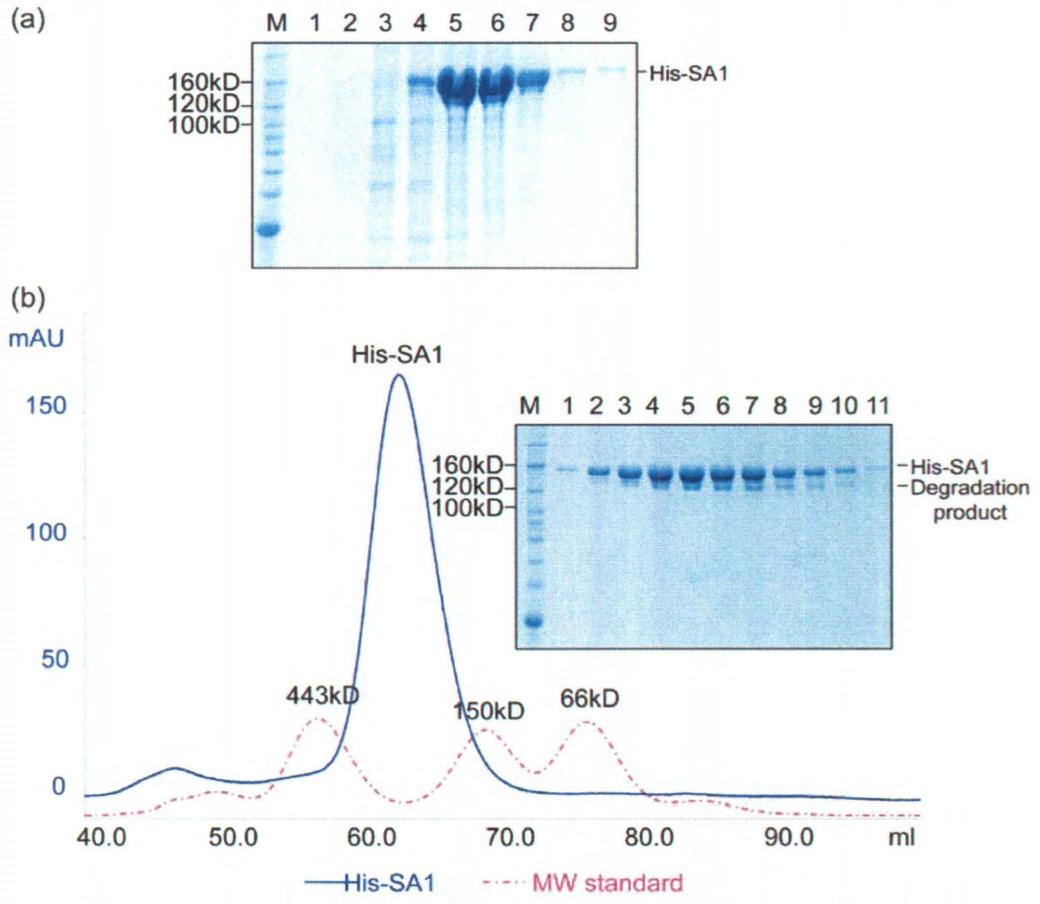


Figure 2.13 Purification of His-SA1 from Sf21 cells. (a) Fractions from the Q column. Lane 1, flowthrough from the Q column; lanes 2-9, fractions eluted from the Q column. Fractions from lanes 5-7 were collected for gel filtration purification. (b) Fractions from the gel filtration column. Lanes 1-10, peak fractions from gel filtration.

```
SA2  1001 ----DILSEF SSKLLRQDKR TVYVYLEKFM TFQMSLRRED  1036
SA1  1001 LAFLEVLSEF SSKLLRQDKK TVHSYLEKFL TEQMMERRED  1040

SA2  1037 VWLPLMSYRN SLLAGGDDDT MSVISG-ISS RGSTVRSKKS  1075
SA1  1041 VWLPLISYRN SLVTGGEDDR MSVNSGSSSS KTSSVRNKKG  1080

SA2  1076 KPSTGKRKVV EGMQLSLTEE SSSSDSMWLS R-EQTLHTPV  1114
SA1  1081 RPPLHKKRVE D--------- -ESLDNTWLN RTDTMIQTPG  1110

SA2  1115 MMQTPQLTST IMREPKR--- ---LRPED-- ----SFMSVY  1142
SA1  1111 PLPAPQLTST VLRENSRPMG DQIQEPESEH GSEPDFLHNP  1150

SA2  1143 PMQTEHHQTP --LDYNRRGT S--------- --------LM  1163
SA1  1151 QMQISWLGQP KLEDLNRKDR TGMNYMKVRT GVRHAVRGLM  1190

SA2  1164 EDDEEPIVED VMMSSEGRIE DLNEGMDFDT MDIDLPPSKN  1203
SA1  1191 EEDAEPIFED VMMSSRSQLE DMNEEFE-DT MVIDLPPSRN  1229

SA2  1204 RRERTELKPD FFDPASIM-D ESVLGVSMF 1231
SA1  1230 RRERAELRPD FFDSAAIIED DSGFGMPMF 1258
```

Figure 2.14 Alignment of the C-terminal sequences of SA2 and SA1. T^{1122} from SA2 is conserved to T^{1123} from SA1. G^{1051} from SA2 is conserved to G^{1055} from SA1. The sequence alignment was prepared using BioEdit (Hall, 1999).

Furthermore, G^{1055} of SA1 is also conserved and equivalent to G^{1051} of SA2. Like SA2, the SA1 1056-1123aa region mainly consists of non-structured coils, which may increase the flexibility of the protein. Therefore, baculovirus overexpressing the SA1 deletion mutants SA1 (1-1123aa) and SA1 (1-1055aa), the latter of which lacks the flexible region, were generated. SA1 (1-1123aa) showed signs of degradation in a small scale purification test (data not shown). Therefore, only the SA1 (1-1055aa) was purified in larger amounts as described in section 2.2 (Figure 2.15). Like SA2 (1-1051aa), SA1 (1-1055aa) was also eluted with an apparent MW slightly greater than 150kD. SA1 (1-1055aa) was then concentrated to 5 mg/ml and crystallization screenings were performed. Unfortunately, no micro crystals were obtained.

2.8 Analytical ultracentrifugation studies of the SA2 protein

Since the crystallization study of the SA2 protein was not successful, we decided to take another approach to obtain structural information for the SA2 proteins. Analytical Ultracentrifugation (AU) was used to address two questions. First, it was mentioned in section 2.3 that SA2 (1-1051aa) was eluted from the gel filtration column at a position indicating an apparent MW larger than 150kD. However, it was previously reported that there was only one SA2 molecule existed in each cohesin complex *in vivo* in both the one-ring model and the handcuff model (Haering *et al.*, 2002; Zhang *et al.*, 2008). By measuring the MW of SA2 using the AU method, we can find out if it forms dimer in insect cell system. Second, the difficulty in crystallization may indicate that the SA2 molecule is largely flexible and its shape is far from being ideal. AU experiments may also provide some information on the shape of the SA2 molecule.

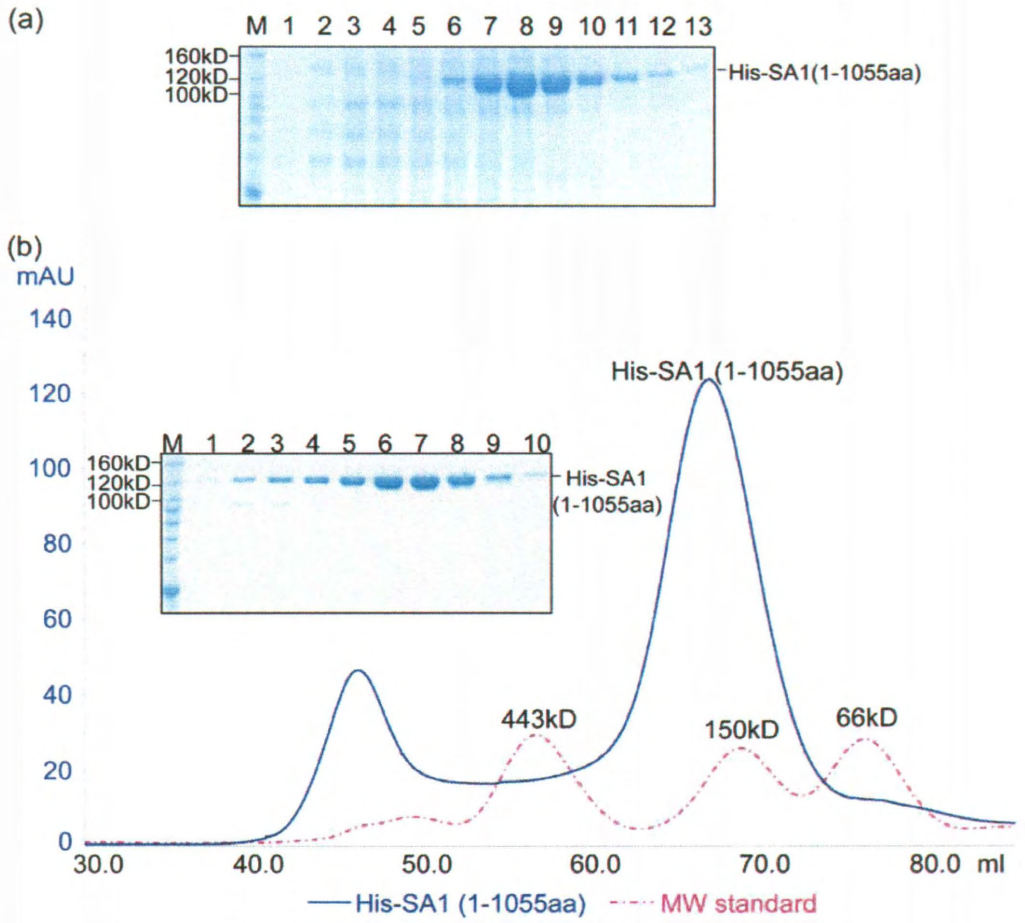


Figure 2.15 Purification of His-SA1 (1-1055aa) from Sf21 cells. (a) Fractions from the Q column. Lane 1, flowthrough from the Q column; lanes 2-13, fractions eluted by the Q column. Fractions from lanes 7-11 were collected for gel filtration purification. (b) Fractions from the gel filtration column. Lane 1, fraction from aggregation peak; lanes 2-10, peak fractions from gel filtration.

AU is a spectroscopic technique, in which a protein sample is spun at a controlled speed and constant temperature. The absorbance of the sample is measured at a given wavelength along the radius of the rotor cells. The concentration distribution of the sample along the cell radius is then determined. There are two AU methods: sedimentation velocity (SV) and sedimentation equilibrium (SE). SV experiments are performed at a relatively higher speed and for a shorter time for the sedimentation of the samples to the bottom of the cells. SE experiments are performed at a relatively lower speed and for a longer time for the samples to reach a concentration distribution equilibrium, in which the force of sedimentation and diffusion acting on the samples are balanced (Demeler, 2010). SE experiments can accurately determine the molecular weights of macromolecules in solution regardless of their shapes. However, the long running time required may cause aggregation and/or the degradation of the protein samples, thus adversely affecting the reliability of the obtained data. SV experiments, when combined with the Ultrascan software as a powerful data analyzing tool (Ultrascan (Demeler, 2005), http: //www.ultrascan.uthscsa.edu), can obtain most of the information traditionally obtained from SE experiments based upon the formulas shown in Figure 2.16. SV experiments also avoid the long running time required for SE experiments. Therefore, SV experiments were used to analyze our samples.

SV experiments were performed for the purified SA2 (1-1051aa) sample using a Beckman Coulter Optima XL-A analytical ultracentrifuge. To determine the MW of SA2 (1-1051aa), data were collected at two different optical density (OD) values, which were obtained against a water reference at 230nm. The sedimentation velocity experiments were performed at 36,000rpm (AN60 Rotor) and 4°C for 15h. One hundred and fifty

$$\left(\frac{\partial C}{\partial t}\right)_r = \frac{-1}{r}\frac{\partial}{\partial r}\left[s\omega^2 r^2 C - Dr\frac{\partial C}{\partial r}\right]_t \quad \text{(a)} \qquad k = \frac{f}{f_o} \quad \text{(b)}$$

$$D(s,k) = RT\left[N\,18\pi\,k\eta\left(\frac{s\bar{v}}{2(1-\bar{v}\rho)}\right)^{1/2}\right]^{-1} \quad \text{(c)}$$

$$M(s,D) = \frac{sRT}{D(1-\bar{v}\rho)} \quad \text{(d)}$$

$$D(f) = \frac{RT}{Nf} \quad \text{(e)}$$

$$s(M,f) = \frac{M(1-\bar{v}\rho)}{Nf} \quad \text{(f)}$$

Figure 2.16 Formulas for the data analysis of sedimentation velocity experiments (Adapted from Demeler, 2005). (a) The relationship between Distribution of the sample concentration (C) along the radius of rotor cells (r) at certain time (t), the sedimentation of the sample (blue circle) and the diffusion of the sample (red circle). S: sedimentation coefficient; ω: angular velocity; D: diffusion coefficient. (b) k is defined as the ratio between frictional ratio of the sample (f) and frictional ratio of the solvent (f_0). (c-f) the relationship between D, M (molecular weight of the sample), k, s and f.

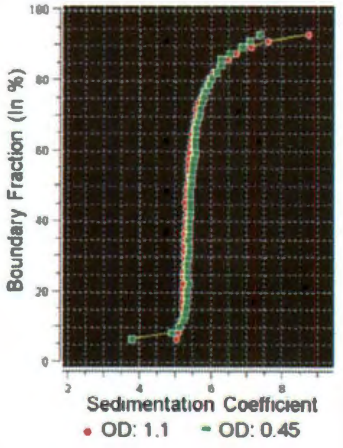


Figure 2.17 G(s) for SA2 (1-1051aa) at different concentrations.

scans were recorded for each sample at a radial step size of 30μm. Data analysis was performed with UltraScan version 9.9 (Ultrascan (Demeler, 2005), http: //www.ultrascan.uthscsa.edu). Time invariant noise and radially invariant noise were subtracted from the sedimentation velocity data by 2-dimentional spectrum analysis (Brookes *et al.*, 2010; www.uslims.uthscsa.edu). The Gaussian distributions of the sedimentation coefficient (G(s)) were obtained with enhanced van Holde-Weischet analysis using Ultrascan (Demeler *et al.*, 2004). The van Holde-Weischet integral distribution plot showed that the distributions of the sedimentation coefficient for both samples are identical (Figure 2.16). This suggests no oligomerization occurred during the data collection. A global generic algorithm Monte Carlo analysis was also performed (Cao *et al.*, 2008) to determine the molecular weights and compositions of the samples. The results showed that there was only one major species displaying a molecular mass of 108.6 (108.2, 109.3) kDa and a frictional ratio of 1.66 (1.65-1.67: 95% confidence intervals). Compared to the theoretical MW of 120kD for monomer and 240kD for dimer, the result showed that SA2 (1-1051aa) was purified from Sf21 cells as monomer, consistent with both of the cohesin models. The frictional ratio of 1.66 suggests that SA2 (1-1051aa) is likely to be an elongated molecule.

2.9 Future direction for the structural studies of the SA2 protein

The AU experiment has answered two questions. First, SA2 (1-1051aa) forms monomer, which is consistent with both the one-ring and handcuff model (Gruber *et al.*, 2003; Zhang *et al.*, 2008). Second, SA2 (1-1051aa) is likely an elongated molecule,

suggesting it might contain one or multiple flexible regions. This might increase the difficulty in crystallization.

To obtain more information of the SA2 protein, we performed a domain prediction using GlobPlot (Linding *et al.*, 2003). The prediction result showed that the N-terminal 75aa of SA2 formed disordered structures, which might be one of the reasons that the crystallization studies of SA2 (1-1051aa) and SA2 (1-1122aa), both of which contained the disordered domain, were not successful. For SA2 (451-1051aa), the N-terminus of this construct did not contain any disordered sequences. However, the C-terminal 982-1051aa contained some disordered sequences. The predicted disordered regions in the SA2 molecular might play a role in increasing the difficulty of crystallization.

In the future, improved schemes can be used in the structural studies of the SA2 protein as described below.

First, globular structural prediction information will be integrated in the design of SA2 mutants. New constructs should exclude the unstructured domains indicated by the globular domain prediction.

Second, after new crystallization targets are identified (i.e., the SA2 mutants that can be purified in large amount), in addition to the crystallization studies, AU experiments will be performed to access the information of the shape of the molecules. For smaller SA2 deletion mutants, circular dichroism (CD) spectroscopy experiments can also be performed to obtain secondary structure information of the molecules.

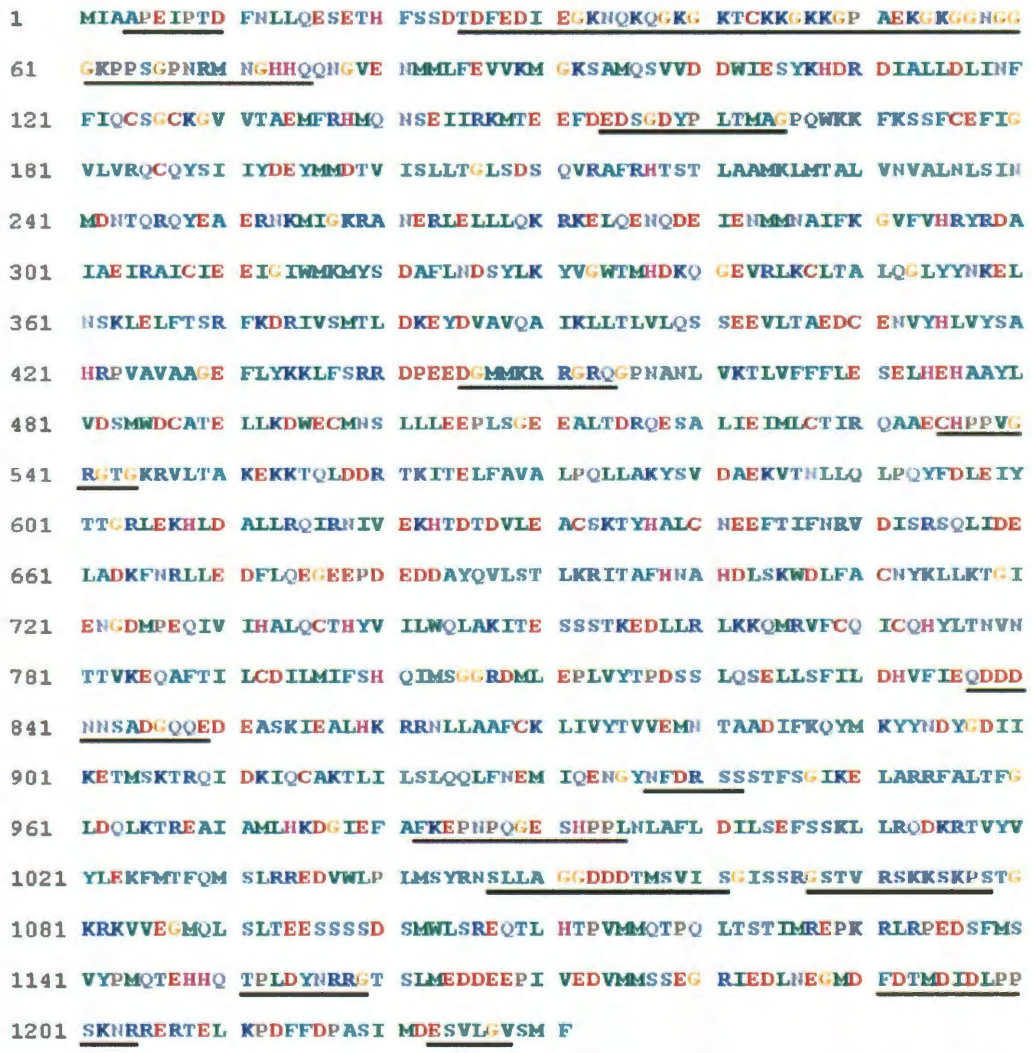


Figure 2.18 Glotplot for the human SA2 protein. Underlined sequences are predicted to be disordered.

Third, because SA2 is known to interact with Rad21 in the cohesin complex (Haering *et al.*, 2002), the interaction domain of SA2 and Rad21 will be determined, and the complex composed of the interaction domains of the two proteins can be used for structural studies. Other than Rad21, Pds5 and Wapl are also shown to interact with SA2 (Rowland *et al.*, 2009). A complex with one of these cohesin associated proteins is also a future direction for the structural characterization of SA2.

Chapter III. The cloning, expression, purification and attempts to study the structure of Rad21

Rad21 connects the NBD heads of Smc1 and Smc3 via its C- and N-terminal domains to form the cohesin ring (Haering *et al.*, 2002; Gruber *et al.*, 2003). It is also the only protein among the three to interact with the fourth cohesin 'core' subunit, SA1/2 (Haering *et al.*, 2002; Gruber *et al.*, 2003). Rad21 is cleaved by a thiol protease, Separase, during the transition from metaphase to anaphase to ensure the complete dissociation of the cohesin complex from sister chromatids (Uhlmann *et al.*, 1999; Hauf *et al.*, 2001). The only known structural information about Rad21, or its homolog, is the C-terminal 115 amino acids of the yeast Rad21, which forms a complex with the NBD of Smc1 consisting of Smc1's N- and C-termini connected by a peptide linker (Haering *et al.*, 2004). Studying the structure of human Rad21 will provide more information on the formation of the cohesin ring and its interaction with other subunits. It will also provide information on whether two Rad21 molecules interact with each other in an anti-parallel manner according to the previously proposed handcuff model (Zhang *et al.*, 2008).

3.1 The cloning and expression of Rad21 and its deletion mutants in insect cells

The Flag-tagged Rad21 gene was cut from 5μg of pFlag CMV2 Rad21 plasmid (courtesy of Dr. Debananda Pati) with SacI and KpnI and ligated into the pFastBac vector. Baculovirus overexpressing the Flag tagged Rad21 was then obtained as described in section 2.1. The cell debris from the P1 virus was analyzed by both Coomassie stained

SDS-PAGE and western blot to determine the expression level of the Rad21 protein. The expression of Rad21 was only detected by western blot but not on a Coomassie stained gel. Therefore, the amount obtained was insufficient for crystallization studies.

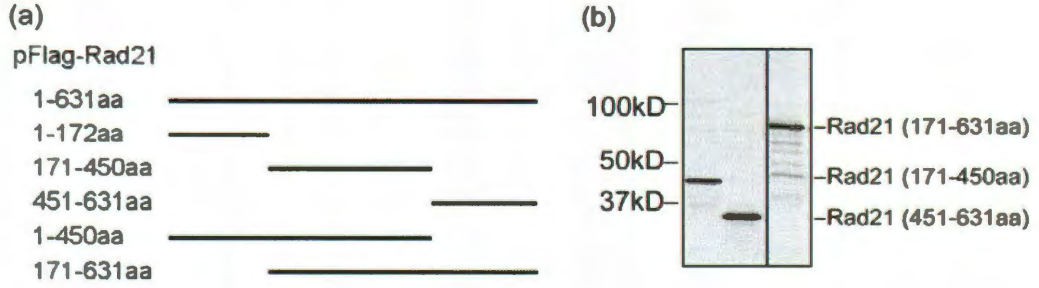


Figure 3.1 Construction and expression of the Rad21 deletion mutants. (a) Schematic figure of the Rad21 deletion mutants. (b) Western blot for Rad21 (171-450aa), Rad21 (451-631aa) and Rad21 (171-631aa). Rad21 (171-631aa) showed signs of degradation during expression.

To identify a Rad21 deletion mutant that was suitable for crystallization studies, five Rad21 deletion constructs were designed based upon the Separase cleavage sites located at R^{172} and R^{450} in human Rad21 (Hauf *et al.*, 2001; Figure 3.1a). Baculoviruses overexpressing each of the deletion mutants with a Flag tag at the N-terminus were generated and the expression levels were tested for each mutant. Among the five deletion mutants, Rad21 (1-172aa) and Rad21 (1-450aa) were only detectable by western blot, while Rad21 (171-450aa), Rad21 (451-631aa) and Rad21 (171-631aa) could be clearly visualized on Coomassie stained gels.

Among the three Rad21 mutants with high expression levels, Rad21 (171-631aa) showed signs of degradation during expression (Figure 3.1b), indicating that it is not a stable protein, and that it is not suitable for crystallization. Therefore, only Rad21 (171-450aa) and Rad21 (451-631aa) were tested for purification. In addition, Flag tagged Rad21 (280-450aa) was also constructed based upon the apoptosis cleavage site at D^{279} (Pati *et al.*, 2002).

3.2 Molecular weights of Rad21 and its deletion mutants

The theoretical MW of human Rad21 is 70kD, but it migrated on a denaturing SDS-PAGE gel with an apparent MW of ~120kD. The Rad21 proteins from other organisms also have a much larger MW than predicted (Guacci *et al.*, 1997). The observed MWs of the Rad21 deletion mutants were also found to be larger than their theoretical MWs at similar proportions (Table 3.1). The cause of the slow migration of Rad21 and its deletion mutants on an SDS-PAGE gel is unknown.

3.3 Purification of the Rad21 deletion mutants from Sf21 cells

Ammonium sulfate precipitation is a cheaper and more convenient method to purify Flag tagged Rad21 compared to using anti-Flag M2 affinity resin (agarose beads conjugated with anti-Flag monoclonal antibody). The three Rad21 deletion mutants that expressed well were first used for a small scale ammonium sulfate precipitation gradient test to determine the amount of ammonium sulfate needed to precipitate the proteins. From the test, I found that Rad21 (171-450aa) precipitated along with other proteins throughout a wide range of ammonium sulfate concentrations. Therefore, ammonium

sulfate precipitation was not a suitable method for the purification of this particular construct. Both Rad21 (451-631aa) and Rad21 (280-450) precipitated at 20% ~ 30% (w/v) ammonium sulfate concentration. Each of these two deletion mutants was purified subsequently.

Rad21 construct	Calculated MW (MW_{cal}; kD)	Apparent MW (MW_{app}; kD)	MW_{app}/MW_{cal}
Full length (631aa)	72.9	120	1.65
1-450aa	52.3	80	1.53
1-172aa	21.0	30	1.43
171-631aa	53.4	90	1.69
171-450aa	32.8	50	1.52
451-631aa	21.9	40	1.83

Table 3.1 Comparison between the apparent MW and calculated MW of Rad21 and its deletion mutants.

Rad21 (451-631aa) was expressed in approximately 1 x 10^9 Sf21 cells. The theoretical pI for Rad21 (451-631aa) is 4.6. Therefore, a Hitrap Q column, which is an anion exchange column, was used as a following purification step using the Rad21 (451-631aa) sample collected from ammonium sulfate precipitation. Rad21 (451-631aa) was eluted from the Q column at 350mM NaCl (Figure 3.2a). The resulting fractions were collected, concentrated, and diluted in FPLC buffer A to reduce the salt concentration to about 100mM NaCl before the sample was loaded onto a 5ml Heparin column. A broad peak was observed in the Heparin chromatogram together with some unbound Rad21

(451-631aa), indicating the affinity of Rad21 (451-631aa) bound to the heparin column was weak (Figure 3.2b). Fractions from the heparin column were then collected, concentrated, and loaded onto a Superdex 200 column. Fractions from the Superdex 200 column were analyzed by Coomassie blue stained SDS-PAGE (Figure 3.2c). Judging from the intensities of the protein bands on the Coomassie blue stained gel, there were two 'peaks' (Figure 3.2c, lane 4 and lane 11). The first peak, indicated by the left arrow in Figure 3.2c, contained aggregated proteins. The second peak, indicated by the middle arrow in Figure 3.2, contained a protein whose apparent MW was around 100kD. The theoretical MW of Rad21 (451-631aa) is ~20kD, suggesting the Rad21 (451-631aa) protein existed as an oligomer in the second peak. The gel filtration chromatogram also had a third peak, indicated by the right arrow in Figure 3.2c, whose suggested MW was smaller than 66kD. However, the proteins from this peak had very low concentrations, as shown by Coomassie gel (Figure 3.2c, lane 17), and the proteins represented may not be the Rad21 (451-631aa) monomer. Indeed, they likely come from the 'shoulder' of the second peak, as the shape of the third peak appeared to be too abrupt for a normal protein. Regardless of the third peak represents the monomer or not, the amount of protein present in either peak two or peak three was insufficient for crystallization studies.

Rad21 (280-450aa) (pI = 5.1) was also expressed in ~1 x 10^9 Sf21 cells. After the Rad21 (280-450aa) sample was precipitated by ammonium sulfate, it was loaded onto a 5ml Q column, from which it was eluted at 400mM NaCl (Figure 3.3a). Fractions containing Rad21 (280-450aa) were loaded onto a superdex200 column and the subsequent fractions obtained were analyzed by Coomassie blue stained SDS-PAGE. Both the gel filtration chromatogram and the Coomassie stained gel showed that most of

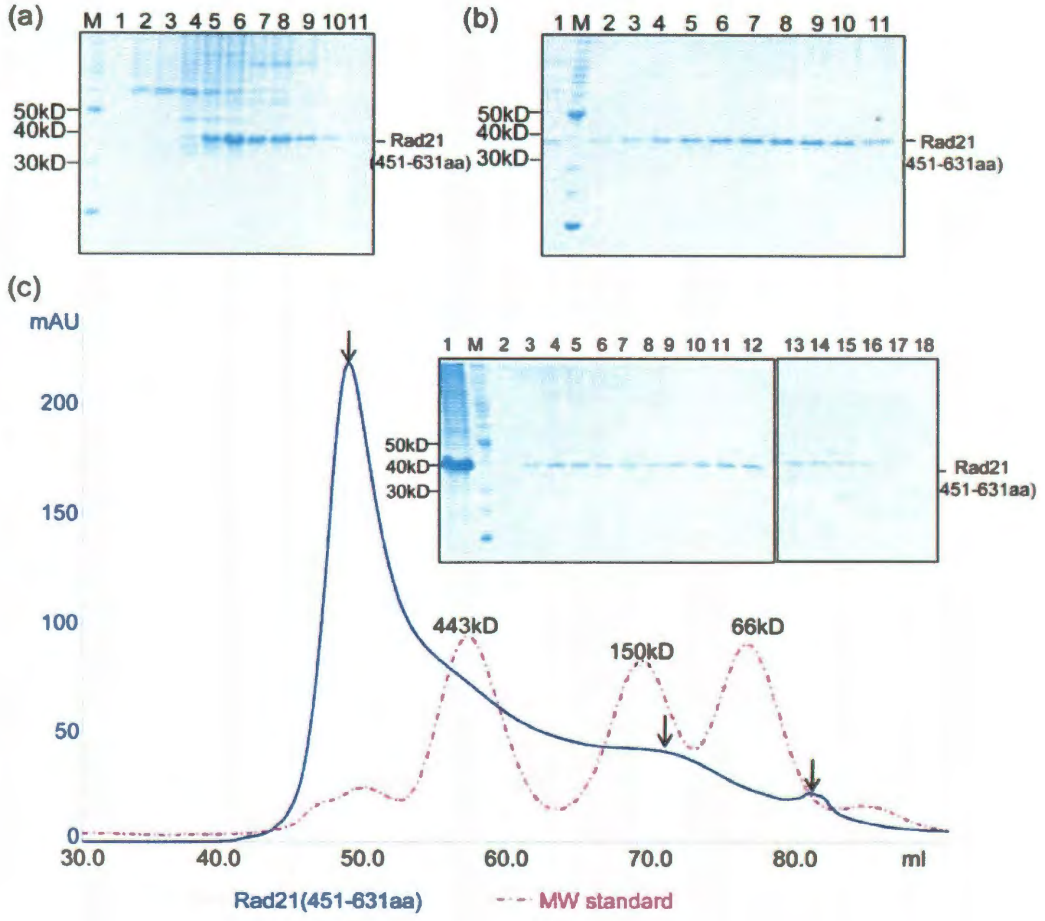


Figure 3.2 Purification of Flag tagged Rad21 (451-631aa) from insect cells. (a) Fractions from the Q column. Lane 1, flowthrough from the Q column; lanes 2-11, fractions eluted by the Q column. Fractions from lanes 6-9 were collected for the heparin column. (b) Fractions from the heparin column. Lane 1, flowthrough from the heparin column; lanes 2-11, fractions eluted by the heparin column. Fractions from lanes 4-9 were collected for gel filtration. (c) The gel filtration chromatogram of Rad21 (451-631aa) and eluted fractions from the column. Lane 1, concentrated samples before being loaded onto the gel filtration column; lanes 2-18, fractions from the gel filtration column.

the Rad21 (280-450aa) was eluted as aggregates (Figure 3.3b left arrow and lane 1). There was also a very small amount of the sample that was eluted at a position indicated by the right arrow in Figure 3.3b, whose apparent MW was less than 66kD. This could correspond to Rad21 (280-450aa) monomer or dimer. However, the amount was insufficient for crystallization studies.

Ammonium sulfate precipitation was not an effective initial step to purify the Flag tagged Rad21 (171-450aa). To aid in the purification of Rad21 (171-450aa), a new Rad21 (171-450aa) construct with a 6 x His tag at the N-terminus was engineered for the baculovirus system. To express the 6 x His tagged Rad21 (171-450aa) (pI=4.4), ~1 x 10^9 Sf21 cells were infected with baculovirus overexpressing Rad21 (171-450aa). Cells were harvested 60h post-transfection and the supernatants obtained from the cell lysate were passed through 1ml of Ni-NTA resin to purify the 6 x His tagged Rad21(171-450aa) (Figure 3.4a, lanes 1-6). In the fractions containing Rad21 (171-450aa), two upper bands were also present. Therefore, the fractions containing the Rad21 (171-450aa) were loaded onto a 5ml Q column, and the protein was eluted at 400mM NaCl. Fractions eluted from the Q column containing Rad21 (171-450aa) were then loaded onto a 1ml Heparin column. Neither the Q column nor the Heparin column was able to separate Rad21 (171-450aa) from the two upper bands. The fractions from the Heparin column were loaded onto a Superose 6 column. However, most of the protein was eluted as aggregates (Figure 3.4c, left arrow). In addition to the large aggregation peak, there was a small peak, indicated by the right arrow in Figure 3.4c, showing that the protein was eluted with an apparent MW larger than 150kD. Given that Rad21 (171-450aa) has a theoretical MW of

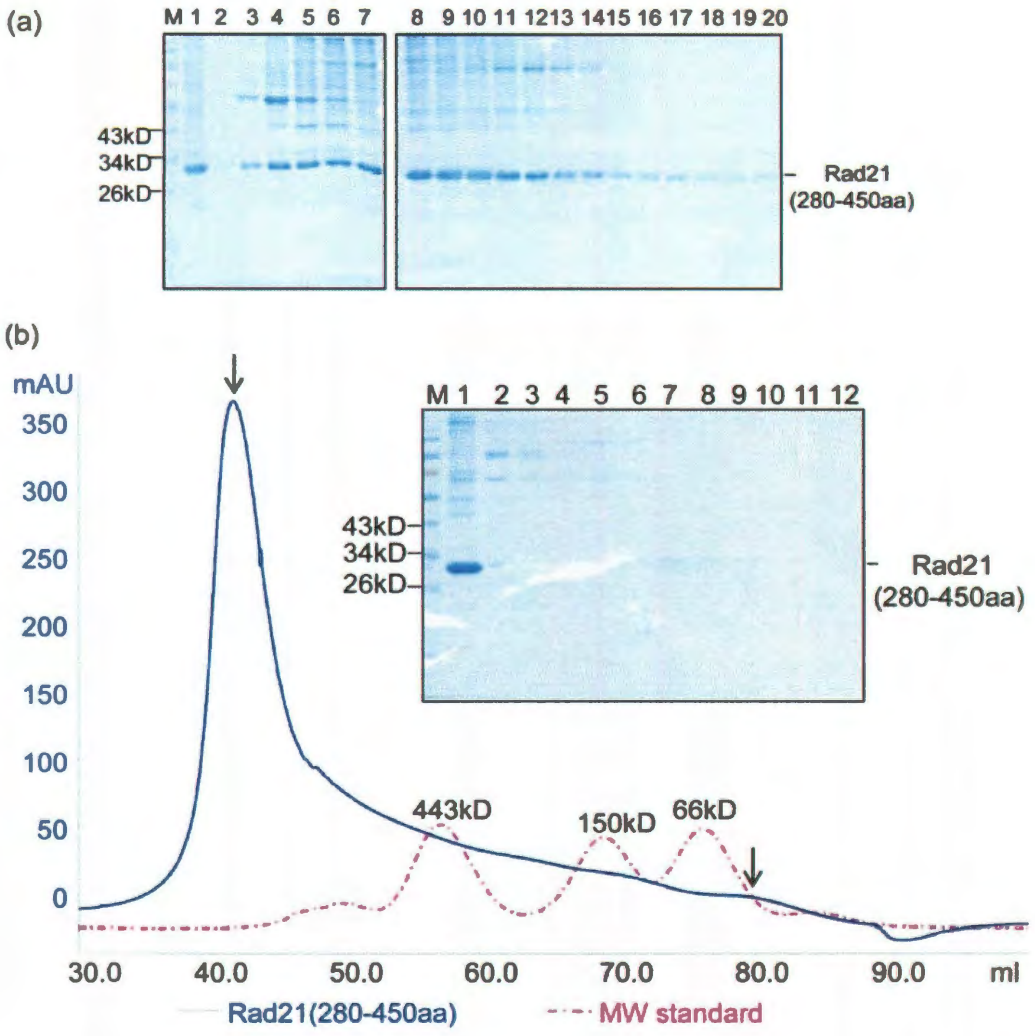


Figure 3.3 Purification of Flag tagged Rad21 (280-450aa) from insect cells. (a) Fractions from the Q column. Lane 1, flowthrough from the Q column; lanes 2-13, fractions eluted from the Q column. Fractions from lanes 8-12 were collected for the heparin column. (b) Gel filtration chromatogram for Rad21 (280-450aa) and fractions from the gel filtration column. Lane 1: fraction from the aggregation peak (left arrow). Lane 2-12, consecutive fractions collected after the aggregation peak. Lanes 6-8 were fractions collected around the right arrow.

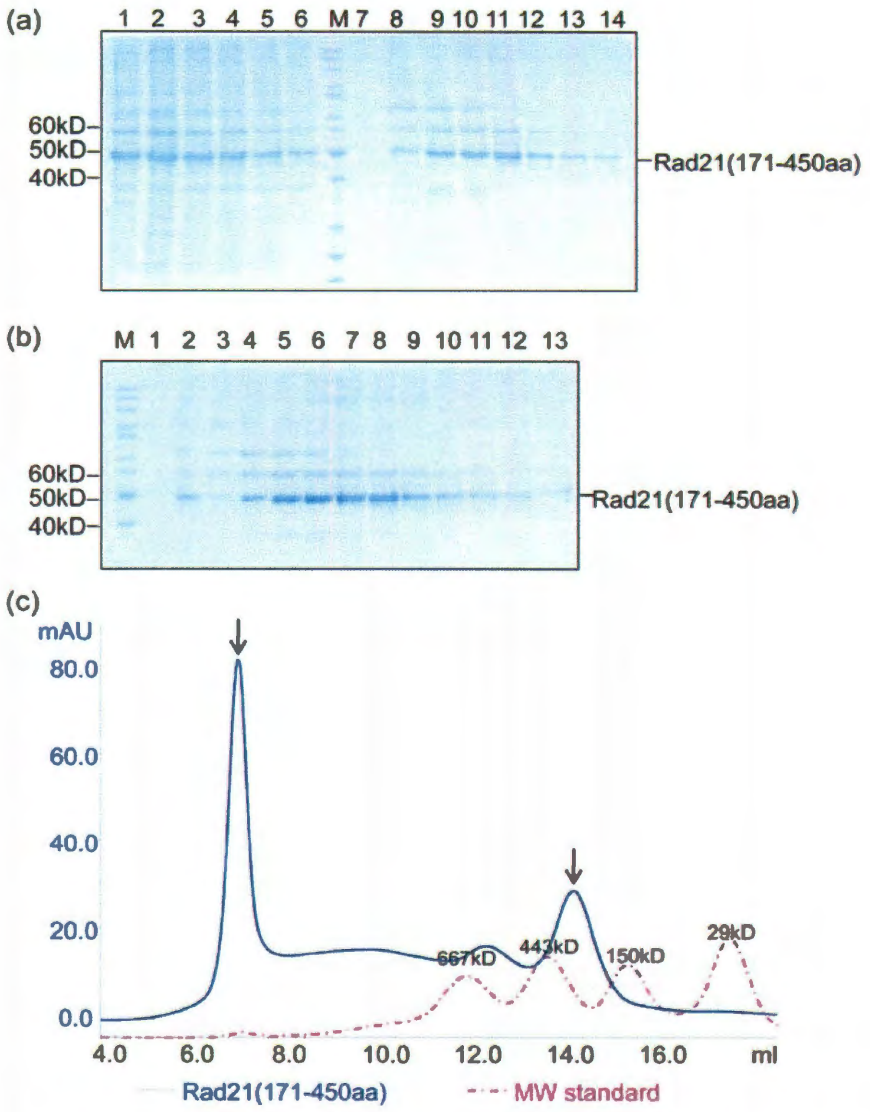


Figure 3.4 Purification of Rad21 (171-450aa). (a) Fractions from the Ni-NTA column and the Q column. Lanes 1-6, fractions eluted by the Ni-NTA column; lane 7, flowthrough from the Q column; lanes 8-14, fractions eluted by the Q column. Lanes 9-11 were collected for the heparin column. (b) Fractions from the heparin column. Lane 1, flowthrough from the heparin column; lanes 2-13, fractions eluted by the heparin column. Lanes 4-9 were collected for gel filtration. (c) Gel filtration chromatogram for Rad21 (171-450aa).

33kD, this peak may represent an oligomer of the protein. However, the amount of this oligomer present was too little for crystallization studies.

The full length Rad21 has been shown to be a 'sticky' protein. When it was co-purified with SA2 (451-1051aa), the complex was eluted mostly as aggregates. SA2 (451-1051aa) can be purified alone without it forming large aggregates (Figure 2.11). Therefore, it is likely that the aggregation of Rad21 caused the aggregation of the complex. Indeed, all three Rad21 deletion mutants that I have tested, including Rad21 (451-631aa), Rad21 (171-450aa), and Rad21 (280-450aa) mostly form aggregates during purification. *In vivo*, Rad21 should not form large aggregates, as unregulated aggregation of Rad21 would result in the aggregation of the cohesin complex and thus lead to the aggregation of the chromosomes, which would be detrimental to cell survival. There are a few possible causes for the nonspecific aggregation of human Rad21 in insect cells. First of all, the expression of Rad21 and its binding to the cohesin complex *in vivo* may be regulated by other human proteins that are missing from insect cells. Another possibility could be that some proteins in the insect expression system may cause the aggregation of Rad21 proteins by an unknown mechanism. The *E. coli* expression system was then used as described below to test this possibility.

3.4 The cloning and expression of Rad21 deletion mutants in *E. coli*

Compared to insect cells, the *E. coli* expression system is less costly and labor intensive. However, it does not provide the post-translational modifications in eukaryotic cells. Therefore, *E. coli* was used to express some of the shorter Rad21 deletion constructs, the folding pathway of which might not be as elaborate the longer Rad21 constructs. 6xHis tagged Rad21 (451-631aa) and Rad21 (280-450aa) were cloned into the

pET28b vector and transformed into Rosetta competent cells. The temperature, time of induction, and IPTG concentration were then varied to determine the optimal expression of each protein.

Rad21 (451-631aa) was optimally expressed at 30°C with 0.1mM IPTG for 4h. Because the Rad21 (451-631aa) construct carries a 6xHis tag at the N-terminus, Ni-NTA resin was used for the initial purification step. However, most of Rad21 (451-631aa) did not bind to Ni-NTA resin, indicating that the N-terminal 6xHis tag might be "buried" in the molecule (Figure 3.5a). In addition, there was a ~70kD protein present in each fragment eluted from the Ni-NTA resin. The ~70kD band was not detected by an anti-6 x His mAb or anti Rad21 mAb (Sigma-Aldrich), suggesting that it was an *E. coli* protein, possibly GroEL, a host chaperonin that is often associated with poorly folded proteins. Alternatively, it could be a cellular protein with a strong Ni-NTA binding affinity. This protein was also present in the same fractions as Rad21 (451-631aa) when they were eluted from a Q column. The amount of the unknown protein is much larger than that of Rad21 (451-631aa) (Figure 3.5b). The 6 x His tagged Rad21 (280-450aa) was also tested using the same method. Similar to Rad21 (451-631aa), the binding affinity of 6 x His tagged Rad21 (280-450aa) to the Ni-NTA beads was low and once again the ~70kD band was dominant in the fractions obtained from the Q column. Given the low efficacy of Ni-NTA purification under natural condition, alternative approaches were required to be tested in order to purify the Rad21 deletion mutants expressed *E. coli*.

For both Rad21 (451-631aa) and Rad21 (280-450aa), around 40%~50% of the expressed proteins were found in inclusion body after cell lysis. Therefore, the proteins

could potentially be purified from the inclusion bodies under denaturing condition and then refolded to their native state.

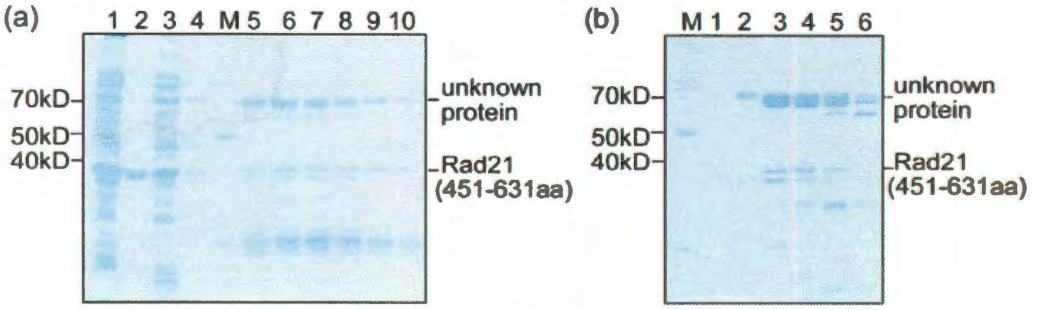


Figure 3.5 Purification of Rad21 (451-631aa) from *E. coli* under native condition. (a) Fractions from the Ni-NTA resin. Lane 1, supernatant; lane 2, cell pellet; lane 3, flowthrough from the Ni-NTA column; lane 4, wash; lanes 5-10, fractions eluted from the Ni-NTA column. (b) Fractions from the Q column. Lane 1, flowthrough from the Q column; lanes 2-6, fractions from the Q column.

There are two different ways to prepare the solution containing the denatured protein. The first way is to lyse the cells directly in 8M urea and collect the supernatants for Ni-NTA purification. The alternative way is to lyse the cells in Ni-NTA buffer A and resuspend the pellets in 8M urea for Ni-NTA purification. The first method is easier to perform while the second method might generate higher purity of sample. Both methods were individually tested using 200ml of *E. coli* cells overexpressing Rad21 (451-631aa). Two solutions containing the denatured proteins, each prepared using one of the above two mentioned methods were loaded onto 1ml Ni-NTA resin respectively before being mixed at 4°C for 1h. Ni-NTA resin was collected by centrifuging the samples at 1000g x

5min. The resin was then washed with wash buffer (50mM Tris-HCl, 8M Urea, pH=6.3). This step was to remove any nonspecifically bound proteins. The bound proteins were then eluted using elution buffer 1 (50mM Tris-HCl, 8M Urea, pH=5.9) and elution buffer 2 (50mM Tris-HCl, 8M Urea, pH=4.5). All of the fractions obtained were analyzed by 12% SDS-PAGE gels. The second method yielded Rad21 (451-631aa) at a much higher purity than the first method (Figure 3.6). Rad21 (451-631aa) was eluted mostly by elution buffer 2 for both methods, indicating a high binding affinity of the denatured protein to the Ni-NTA beads.

With the above encouraging results, I then performed Rad21 purification at a large scale. Rad21 (451-631aa) was expressed in approximately 600ml of *E. coli* cells and purified using the second method as described above. The fractions containing the purified Rad21 (451-631aa) showed ~95% purity which is sufficient for gel filtration purification. However, it needed to be refolded before being loaded onto the superdex200 column. The protein was refolded using a refolding buffers (pH=8.2) containing 50mM Tris-HCl, 20mM NaCl, 0.8mM KCl and varying amount of guanidine and arginine. A large amount of precipitation was observed during the course of refolding under each condition. Among these, the condition containing 0.4M arginine generated the least amount of precipitation. After refolding, the Rad21 (451-631aa) sample acquired additional band on SDS-PAGE gel at the ~110kD position. This was approximately three times as large as the MW of the Rad21 (451-631aa) monomer, suggesting the formation of a trimer (Figure 3.7b, lane1 vs. lane2), the cause of which is unknown. The precipitation was removed by a 35,000g x 30min centrifugation and the supernatant was loaded onto the Superdex 200 column. The fractions eluted from the Superdex 200

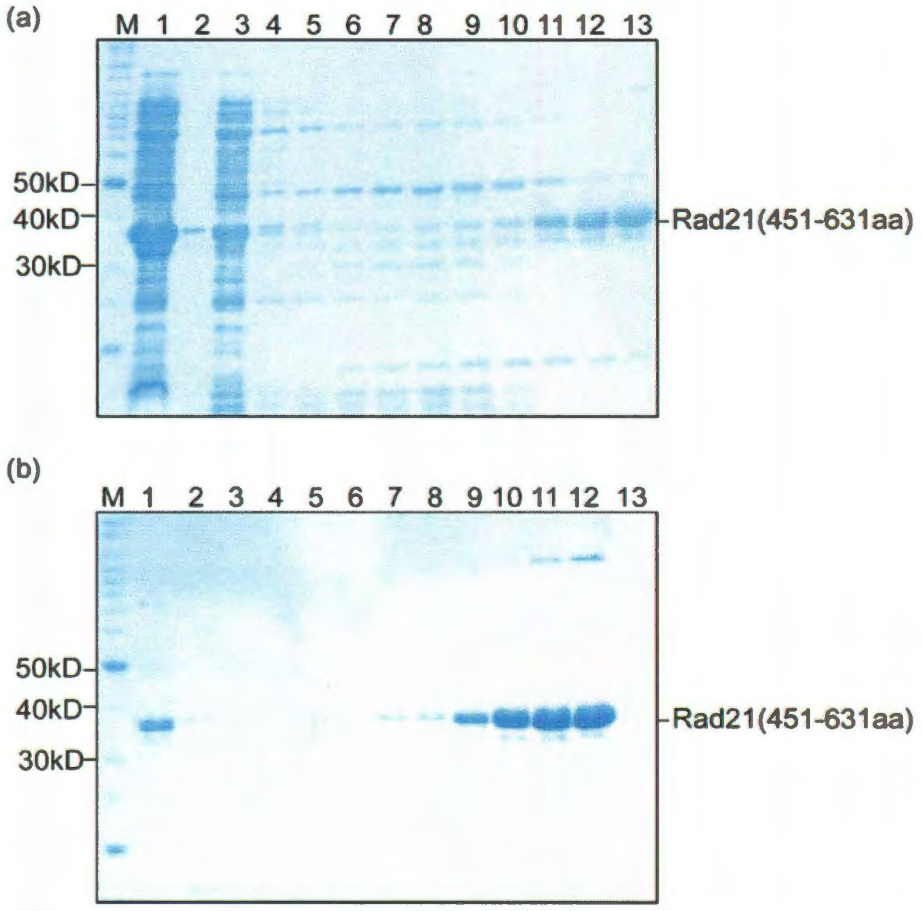


Figure 3.6 Purification test of 6 x His tagged Rad21 (451-631aa) using Ni-NTA resin under denaturing condition. (a) Ni-NTA purification using the first method as mentioned in section 3.4. Lane 1, supernatant; lane 2, cell pellet; lane 3, flowthrough from the Ni-NTA column; lanes 4-6, wash; lanes 7-10, fractions eluted by elution buffer 1; lanes 11-13, fractions eluted by elution buffer 2. (b) Ni-NTA purification using the second method as mentioned in section 3.4. Lane 1, flowthrough; lanes 2-4, wash; lanes 5-8, fractions eluted by elution buffer 1; lanes 9-13, fraction eluted by elution buffer 2.

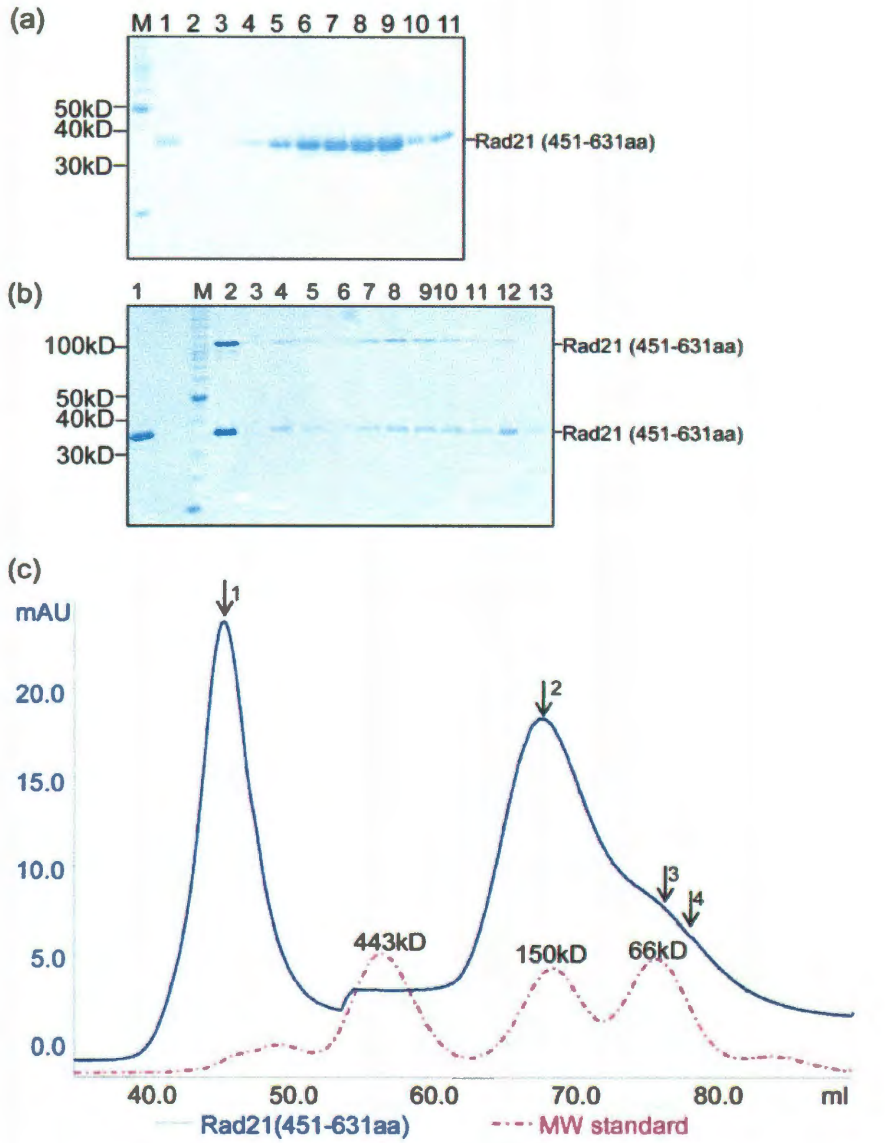


Figure 3.7 Purification of refolded Rad21 (451-631aa). (a) Fractions purified by Ni-NTA column under denaturing condition. Lane 1, flowthrough; lanes 2-3, wash; lane 4, fraction eluted by elution buffer 1; lanes 5-11, fraction eluted by elution buffer 2. (b) Fractions from the gel filtration column. Lane1, samples before refolding; lane 2, samples after refolding and before being loaded onto the gel filtration column; lanes 3-13, fractions eluted from the gel filtration column. (c) The gel filtration chromatogram for Rad21 (451-631aa).

column were analyzed by a 12% SDS-PAGE gel and subsequently stained by Coomassie blue. All the fractions analyzed contained the upper band except for lane 13 (Figure 3.7b). Three 'peaks' could be assigned to the gel filtration chromatogram based upon the protein concentration in different fractions on the Coomassie stained gel (Figure 3.7b lanes 4, 8, and 12). The three peaks were indicated by arrows 1, 2, and 3 in Figure 3.7c. The first peak contained only aggregates. The MW indicated by the second peak was approximately 150kD, suggesting that it contained both cross linked Rad21 (451-631aa) trimer and other Rad21 (451-631aa) oligomers. There was only one fraction that contained the lower band and not the upper band (Figure 3.7b, Lane13; Figure 3.7c, arrow 4). However, the amount obtained was too small for crystallization studies.

To purify the 6 x His tagged Rad21 (280-450aa), a similar approach as described above was used. Rad21 (280-450aa) was overexpressed in 1L of *E. coli* and eluted at pH 4.5 from the Ni-NTA resin. To refold the protein, the purified fractions were diluted twenty times in a refolding buffer containing 0.4M arginine. The protein sample was added to the dilution buffer in a drop wise manner with constant mixing. The mixture was then dialyzed against a fresh dialysis buffer (50mM Tris-HCl, 20% Glycerol, 500mM NaCl, and 1mM EDTA) for 3 x 6h. Similar to Rad21 (451-631aa), precipitates were observed during the refolding process and part of the refolded protein also formed a trimer that was difficult to dissociate (Figure 3.8b). A refolding buffer containing 0.4M arginine and an additional 2mM GSH/0.4mM GSSG was also tested using the same procedure as described above. The addition of GSH/GSSG significantly increased the recovery rate of the protein sample (Figure 3.9b), but the refolded product still contained the cross linked trimer. Upon analysis of the eluted samples, it was found that most of the

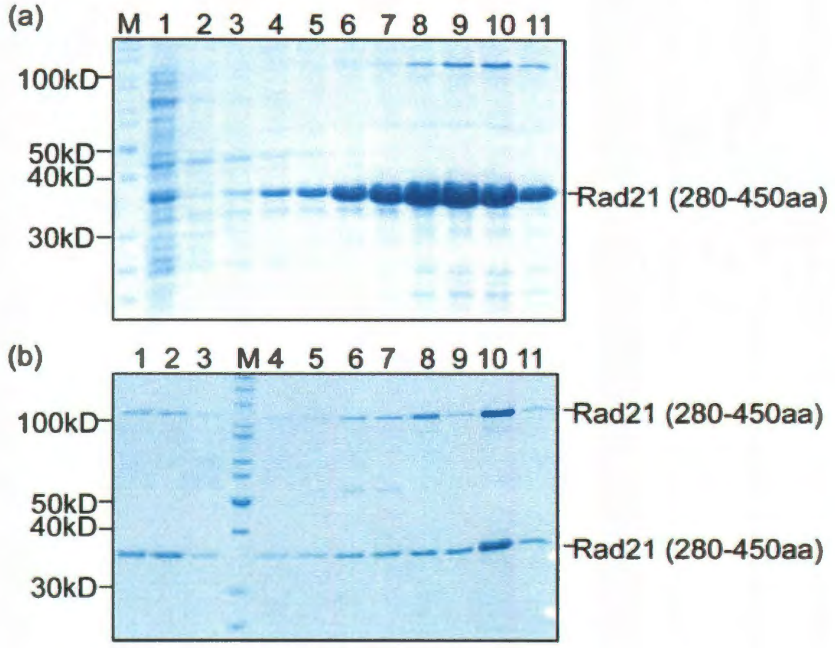


Figure 3.8 Purification of Rad21 (280-450aa) from E. coli. (a) Fractions purified by Ni-NTA resin under denaturing condition. Lane 1, flowthrough; lanes 2-3, wash; lane 4, fraction eluted by elution buffer 1; lanes 5-11, fraction eluted by elution buffer 2. (b) Fractions from the gel filtration column using refolded proteins. Lanes 1-3, fractions collected from the aggregation peak; lanes 4-11, fractions collected from the non-aggregation peak.

proteins were eluted with an apparent MW of ~150kD. The proteins in these fractions were either cross linked trimers or oligomers of Rad21 (280-450aa). There was a small amount of Rad21 (280-450aa) protein, as indicated by the right arrow in Figure 3.9c, that was not associated with cross-linked oligomers. However, a trace amount of the cross linked trimer was still present in these fractions.

The fractions from the third peak (i.e., the peak indicated by the right arrow in Figure 3.9c) could be collected and purified by the gel filtration column a second time, which may result in a more pure protein sample without the contamination of the cross linked trimer. However, a second round of purification by a gel filtration column is likely to cause the loss of some of the protein sample, resulting in a smaller amount of protein in the final sample. Because the Rad21 (280-450aa) protein was purified from *E. coli*, which is cheap and easy to grow, a larger scale expression and purification can be used to obtain sufficient Rad21 (280-450aa) for crystallization studies.

The purification of Rad21 (280-450aa) has brought us hope to obtain sufficient amount of Rad21 mutant protein for crystallization studies. Because the purification of Rad21 (280-450aa) requires the protein to be refolded from denaturing condition, it needs to be tested prior to crystallization screening for correct folding. There is no immediate biophysical assay available to test the correct folding of Rad21 thus far. However, we can indirectly determine whether it is correctly folded by testing one of its biological functions. We have confirmed that Rad21 (280-450aa) can co-purify with SA2 (1-1051aa) in chapter IV. Therefore, correctly folded Rad21 (280-450aa) should be able to form complex with SA2 (1-1051aa) *in vitro*. Section 2.3 has shown that SA2 (1-1051aa) can be purified in large amount at high purity. To confirm the correct folding of Rad21 (280-

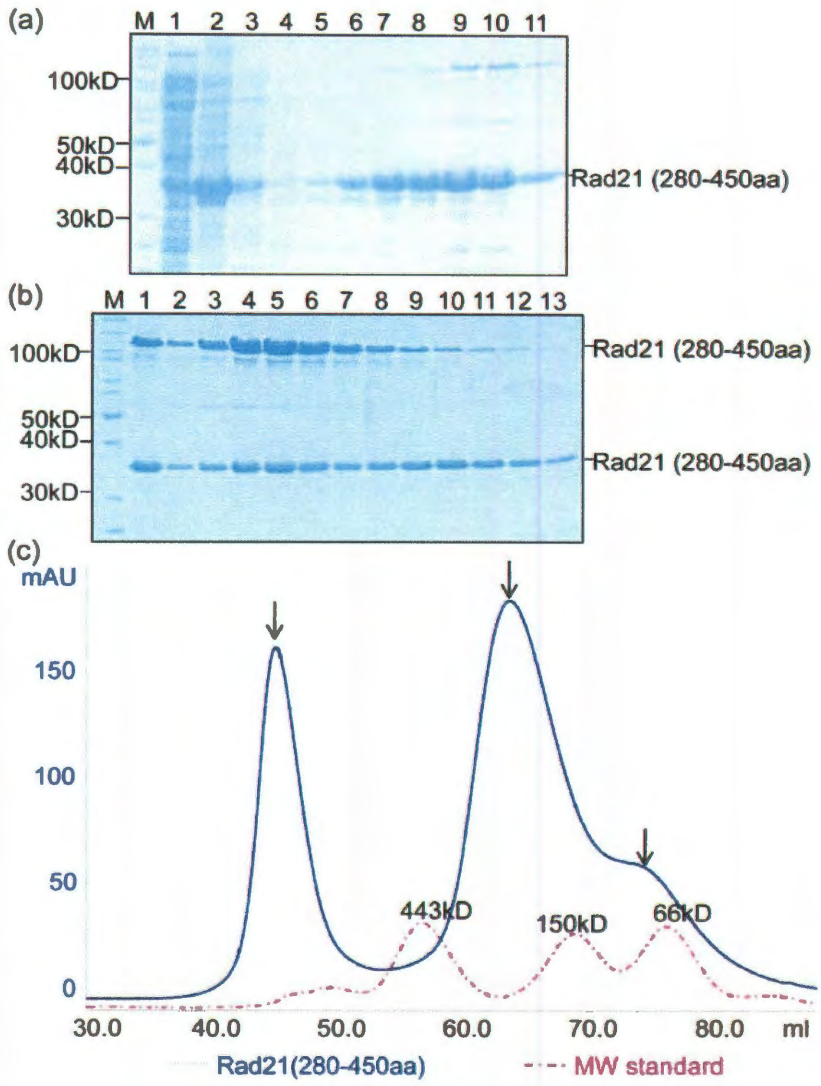


Figure 3.9 Purification of Rad21 (280-450) denatured and refolded using GSH/GSSG. (a) Ni-NTA purification of Rad21 (280-450aa) under denaturing condition. Lane 1, supernatant; lane 2, pellets dissolved in 8M urea; lane 3, flowthrough from Ni-NTA resin; lane 4, wash; lane 5, fraction eluted by elution buffer 1; lanes 6-11, fractions eluted by elution buffer 2. (b) (c) Gel filtration chromatogram (c) and eluted fractions on SDS-PAGE (b). Lane 1, fraction from the aggregation peak (indicated by the left arrow); lanes 2-8, fractions from the second peak (middle arrow); lanes 9-13, fractions from the third peak (right arrow).

450aa), the purified Rad21 (280-450aa) protein will be mixed with purified SA2 (1-1051aa) and the mixture samples will be analyzed by gel filtration column. The eluted peak will be used to compare with the peak of SA2 (1-1051aa) alone to determine whether they form a complex or not. Because the molar ratio of the Rad21 (280-450aa): SA2 (1-1051aa) complex is unknown, different molar ratios (1:1, 1:2, 2:1) will be tested in the gel filtration purification.

3.5 Future direction of structural studies of Rad21

The Rad21 deletion mutants were difficult to purify and readily formed aggregates, suggesting disordered structures of the Rad21 protein. A secondary structure prediction was performed using PredictProtein (Rost *et al.*, 2004) for the human Rad21 (Figure 3.10). The secondary structure prediction showed that the middle region of Rad21 (171-450aa) was largely unstructured / disordered. The region only contained three α-helices, one of which was predicted with very low probability. Therefore, it might be very difficult to crystallize the protein. However, section 3.4 has shown that it is possible to obtain sufficient amount of Rad21 (280-450aa), which is part of the middle region of Rad21. To confirm its predicted secondary structure, purified Rad21 (280-450aa) can be used for CD measurement and secondary structure calculation. The results obtained from CD will be compared to the predicted secondary structure. In addition, nuclear magnetic resonance (NMR) spectroscopy is a feasible way to analyze the structure of the protein due to its small size. Another way to study the structure of the Rad21 middle region is to study the structure of a complex formed by the Rad21 middle region and another Rad21

binding protein. The formation of the complex may decrease the flexibility of the Rad21 molecule and facilitate the formation of crystals.

The secondary structure prediction also showed that, compared to the middle region of Rad21, the N-terminus (1-170aa) and C-terminus (451-631aa) were more structured, containing more α-helices and β-strands with higher probability, which made them better targets for crystallization (Figure 3.10). The Rad21 protein is conserved from yeast to human, sequence alignment showed ~33% similarity between the yeast and the human Rad21 proteins (Figure 3.11). The N-terminus (1-170aa) shares ~40% similarity with the yeast Rad21, while the middle region and the C-terminus shares ~30% similarity. A secondary structure prediction for the yeast Rad21 also showed a largely unstructured middle region of the yeast Rad21 (Figure 3.12) and structured N- and C- termini. Therefore, the crystallization studies of the human Rad21 protein will focus on the N- and the C- termini in the future.

The whole human Rad21 N-terminus has been shown difficult to express in section 3.1. By examining the secondary structure prediction result, the helices and strands all located within the first 90aa (Figure 3.10). In yeast, the N-terminal 1-115aa has been shown to be sufficient binding to Smc3 (Gruber *et al.*, 2003). The human 1-100aa is conserved to the yeast 1-115aa (Figure 3.11), suggesting it might be a structured functional region on its own. This region also excluded the largely unstructured 101-170aa region, which may pose a problem for protein expression and purification. Therefore, the human Rad21 1-100aa can be used as a target for crystallization.

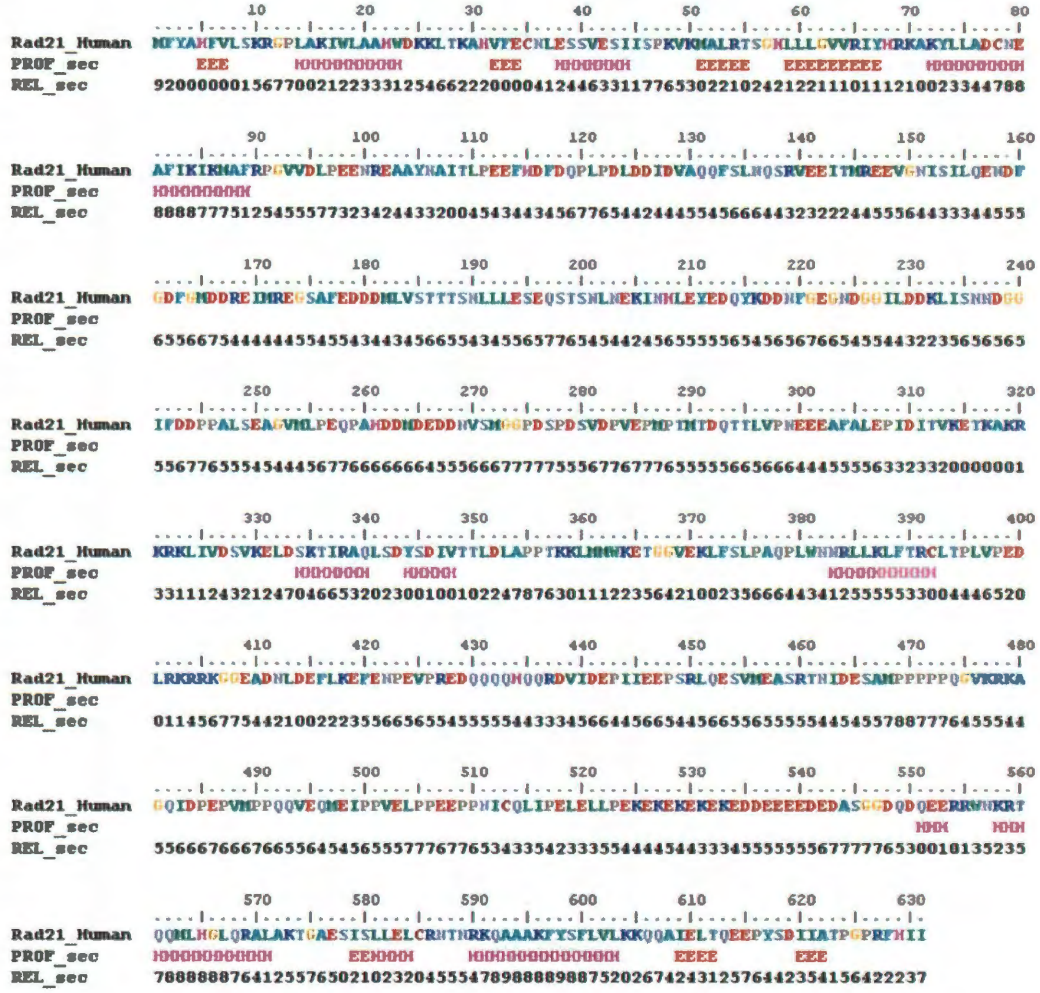


Figure 3.10 Secondary structure prediction for the human Rad21. Rad21 amino acid sequence is colored based on residue types (e.g. blue for positively charged, red for negatively charged, green for hydrophobic, silver for polar, etc). PROF_sec predicts the secondary structure (H = Helix, E = strand). Rel_Sec shows the reliability index of the PROF_sec prediction (0=low, 9=high).

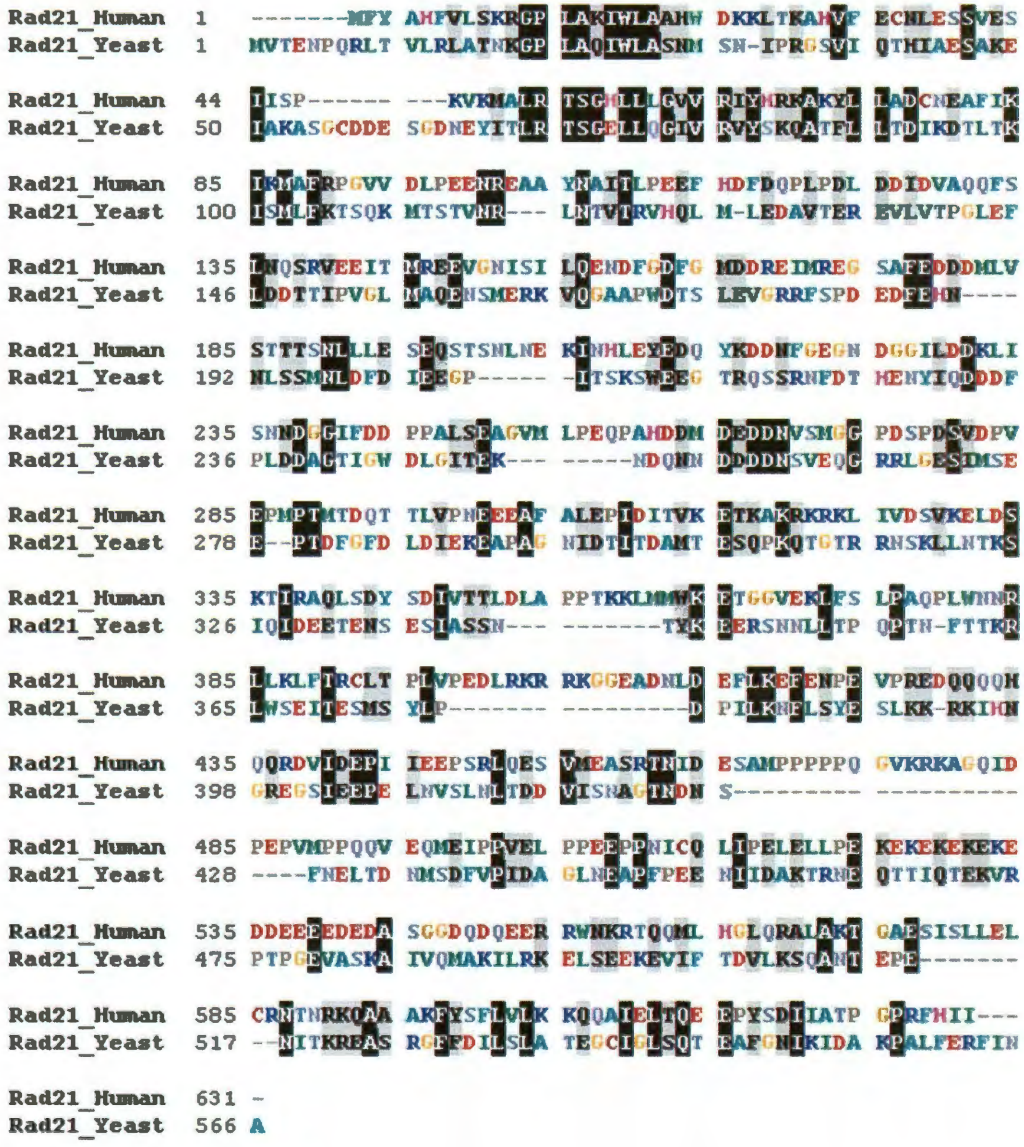


Figure 3.11 Alignment of the human and yeast Rad21. The sequence alignment was prepared using BioEdit (Hall, 1999).

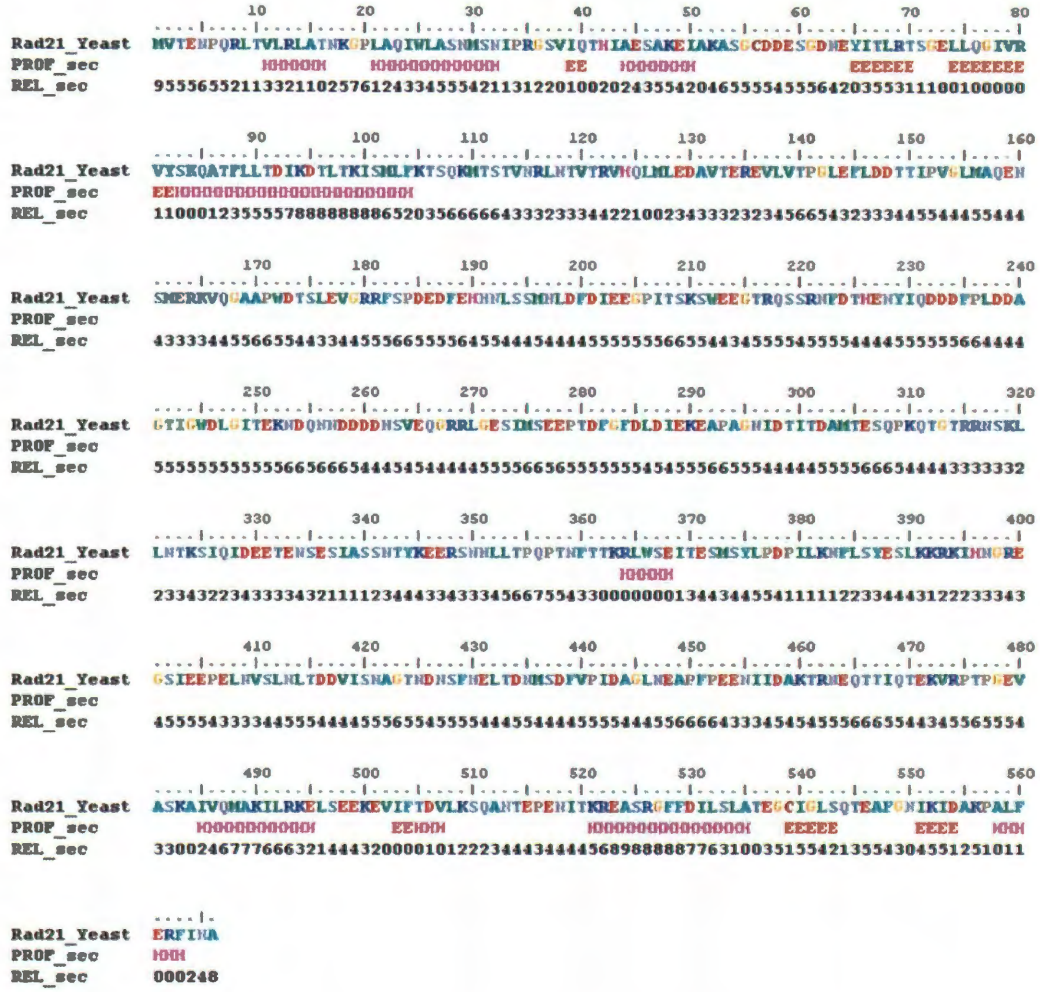


Figure 3.12 Secondary structure prediction for the yeast Rad21. Rad21 amino acid sequence is colored based on residue types (e.g. blue for positively charged, red for negatively charged, green for hydrophobic, silver for polar, etc). PROF_sec predicts the secondary structure (H = Helix, E = strand). Rel_Sec shows the reliability index of the PROF_sec prediction (0=low, 9=high).

In addition to the N-terminus, crystallization targets can also be designed according to the C-terminal structural information. The structure of the yeast C-terminal (451-566aa) has been solved as mainly forming winged helix domains. The 512-631aa from the human Rad21 is conserved to this part and it contains multiple α-helices and β-strands with high probability. The human (451-631aa) has been shown hard to purify, which might be due to the unstructured 451-511aa region. The shorter 512-631aa might be a better crystallization target.

To study the structure of human Rad21, Rad21 (1-100aa) and Rad21 (512-631aa) will be cloned and expressed in *E.coli*. If the proteins are hard to express or purify, insect cell system will be used as an alternative. The yeast Rad21 C-terminal 451-566aa, whose structure was solved, was expressed in insect cell system instead of yeast system (Haering *et al.*, 2004), which might indicate that the Rad21 protein is very difficult to express and purify. The human Rad21 might adopt a more complex folding pathway and more elaborated modifications compared to its yeast homolog. Therefore, a more delicate expression system might be required to prepare the human Rad21 protein.

In addition to the crystallization method, NMR spectroscopy can be an alternative method to determine the structure of the Rad21 mutants due to their small size.

If it is still hard to solve the structure of human Rad21, the Rad21 protein from other mammals can be tested. Rad21 is a highly conserved protein among mammals. Small differences in sequence can have dramatic crystallization of a protein. Rad21 proteins from other mammals, such as mice, can be tested to determine if they are good crystallization targets.

Chapter IV. Biochemical and biophysical characterization of the SA2:Rad21 complex

It has been previously reported that in yeast, Scc3 (known as SA1/2 in human) associates with the cohesin ring via interacting with Scc1 (known as Rad21 in human; Haering *et al.*, 2002). However, it is unclear whether the physical association of human SA2 and Rad21 takes place through direct binding and/or if this association requires additional cohesin associated components for their association in the cohesin ring to take place. Furthermore, a detailed map of the Rad21-SA2 interaction domains and the functional mechanism of SA2 in sister chromatid cohesion have not been described. The studies on the interaction between the human Rad21 and SA2 are reported in this chapter.

4.1 The interaction between Rad21 and the SA2 deletion mutants

To study the interaction between Rad21 and SA2, the interacting regions of both proteins need to be mapped. A co-purification method combined with mutagenesis was used to map the regions of SA2 that can interact with the full length Rad21. The mapping started by examining two SA2 deletion mutants, SA2 (1-1051aa) and SA2 (1052-1231aa), the sequences of which do not overlap with each other. Among them, SA2 (1-1051aa) was identified in section 2.3 as a stable SA2 deletion mutant that did not degrade during the purification process.

To examine the interaction between the Flag tagged Rad21 and the two 6 x His tagged SA2 deletion mutants, ~2 x 10^7 sf21 cells were infected with baculovirus

overexpressing Rad21 along with baculoviruses overexpressing either SA2 (1-1051aa) or SA2 (1052-1231aa) at an MOI of 4:2 respectively. Cells were collected 48h post-infection by centrifugation at 1,000g for 6min. The cell pellets were washed with PBS and resuspended in 2ml of lysis buffer. The resuspended cells were incubated on ice for 30min, with occasional vortexing. After incubation, the supernatants were collected by an 18,000g x 15min centrifugation. The supernatants were split into two halves. 1ml of supernatants from each sample was added to 20µl of Ni-NTA agarose beads supplemented with 5M NaCl and 1M imidazole, to achieve a final concentration of 300mM NaCl and 20mM imidazole. Another 1ml of supernatant from each sample was added to 20ul Flag monoclonal antibody (mAb) conjugated agarose beads (Flag beads; Sigma-Aldrich) supplemented with 5M NaCl, to achieve a final concentration of 150mM NaCl.

For Ni-NTA co-purification, the samples were rotated at 4°C for 1h and washed three times with 340ul of Ni-NTA buffer A and the tagged proteins were then eluted with 2 x SDS sample buffer (125mM Tris-HCl, pH=6.8, 20% glycerol, 4% SDS, 0.1% bromophenol blue, 100mM DTT). For Flag beads co-purification, the samples were gently rotated at 4°C for 3h, followed by three 5min washes with 340ul PBS. The tagged protein was eluted by 2 x SDS sample buffer. The obtained co-purification samples, together with the supernatants, were analyzed by western blot using either an anti-Flag polyclonal antibody (pAb; Sigma-Aldrich) or an anti-6 x His mAb respectively. The influenza A virus PA protein with a 6 x His tag at the N-terminus was used as a negative control because the PA protein is a non-sticky monomer which does not bind to the Rad21 (Guu *et al.*, 2008). To run the western blot, cell lysates and co-purification

samples were loaded onto a 5% - 20% gradient SDS-PAGE gel, and electrophoresed at 175V for about 50min. The samples were then transferred to a nitrocellulose membrane (Bio-Rad) at 110V for 60min. The membranes were blotted with blocking buffer (LI-COR Biosciences, Lincoln, NE) for 1h before being probed with the appropriate primary antibodies for 1h, followed by three washes in TBST buffer (pH=7.4, 20mM Tris-HCl, 150m NaCl, 0.1% Tween 20). The membranes were then probed with appropriate secondary antibodies labeled with IRDye 800 or Cy 5.5 for 1h, followed by three washes in TBST buffer. The membranes were then visualized by an Odyssey Infrared Scanner (LI-COR Biosciences).

The reciprocal Ni-NTA and Flag co-purification results showed that SA2 (1-1051aa) but not the C-terminal SA2 (1052-1231aa) could co-purify with Rad21 (Figure 4.1). The results suggested that the N-terminal 1051aa are sufficient and that the C-terminal 180aa are not required for the Rad21-SA2 interaction.

To further narrow down the region of SA2 responsible for the Rad21-SA2 interaction, the other SA2 deletion mutants obtained in section 2.5 were tested in the co-purification experiment with Rad21. Rad21 was expressed with each of the SA2 deletion mutants (Figure 2.11a) in Sf21 cells and the co-purification procedures using both Ni-NTA and Flag beads were carried out as described above. The co-purified samples were analyzed by western blot. Results showed that amino acids 1-302aa and 751-1051aa are not critical for the SA2-Rad21 interaction (Figure 4.2). It was noticed that, SA2 (1-903aa) showed a weak interaction, while SA2 (1-750aa) and SA2 (1-1051aa) both showed a strong interaction with Rad21 (Figure 4.2). The weak interaction of SA2 (1-903aa) might be a result of a disturbed structure/protein misfolding. Indeed, GlobPlot (Linding *et al.*,

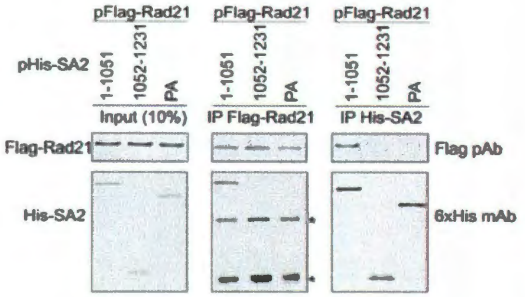


Figure 4.1 Rad21 interacts with SA2 (1-1051aa). Flag tagged Rad21 was co-expressed in Sf21 cells with 6 x His tagged either SA2 (1-1051aa) or SA2 (1052-1231aa) and co-purified using Ni-NTA or Flag beads. Rad21 co-expressed with the influenza A virus PA protein was used as a negative control. Western blot analysis was carried out using either the FLAG polyclonal antibody (pAb) or the 6xHis monoclonal antibody (mAb). Nonspecific bands are marked by asterisks.

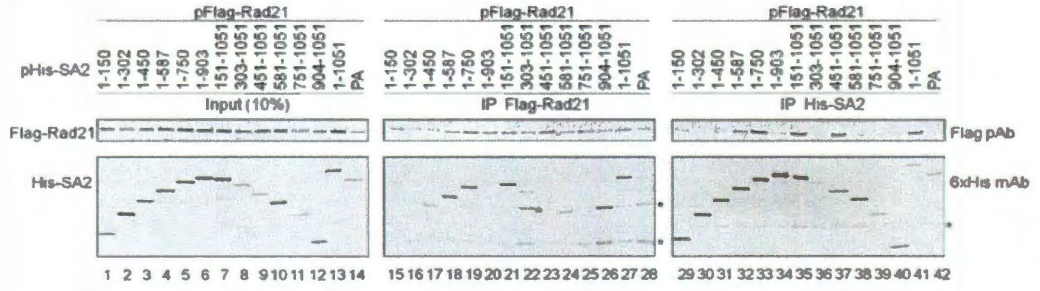


Figure 4.2 Characterization of the interaction between Rad21 and the SA2 mutants. Flag-tagged Rad21 WT was co-expressed with His-tagged SA2 deletion mutants and co-purified by Ni-NTA or Flag beads. The influenza A virus PA was used as a negative control. Nonspecific bands are marked by asterisks.

2003) also predicted that SA2 (850-940aa) might form a globular domain, consistent with our hypothesis. Interestingly, 1-450aa and 587-1051aa, two regions without any overlapping sequences, both interacted with Rad21 (Figure 4.2, lanes 17 & 31 and 24 & 38), indicating that SA2 interacts with Rad21 via multiple peptide stretches and that the structural integrity of the whole SA2 protein, or at least the region containing amino acids 303-750, may be important for the Rad21-SA2 interaction.

4.2 The interaction between SA2 and the Rad21 deletion mutants

Previous studies have shown that the yeast Rad21 is cleaved by Separase at R^{180} and R^{268} (Uhlmann *et al.*, 1999) and the yeast SA2 binds to the Rad21 C-terminal Separase cleavage product (269-566aa) (Haering *et al.*, 2002; Gruber *et al.*, 2003). How human SA2 binds to Rad21 remains unknown. Human Rad21 is cleaved by separase at R^{172} and R^{450} (Hauf *et al.*, 2001). To determine the region of human Rad21 that interacts with SA2, the Rad21 deletion mutants described in section 3.1 were tested (Figure 3.1).

SA2 (1-1051aa), with a 6 x His tag at the N-terminus, was expressed along with the full length Rad21 and each of its deletion mutants, all Flag-tagged at their N-terminus. Co-purification was then performed as described previously using Ni-NTA and Flag beads to examine the interaction between SA2 and the Rad21 deletion mutants. Sf21 cells infected with SA2 (1-1051aa) and the Flag tagged influenza PA protein were used as a negative control. The results showed that SA2 (1-1051aa) can form a complex with Rad21 (171-450aa) but not with Rad21 (1-172aa) or Rad21 (451-631aa). This indicates that unlike in yeast, where the C-terminal Separase-cleaved fragment was responsible for

interacting with Scc3, in humans only the middle Separase cleavage fragment of Rad21 binds to SA2 (Figure 4.3). SA2 (1-1051aa) seemed to be able to pull down a small amount of Rad21 (451-631aa) (Figure 4.3), which was likely due to a non-specific interaction and the high expression level of Rad21 (451-631aa).

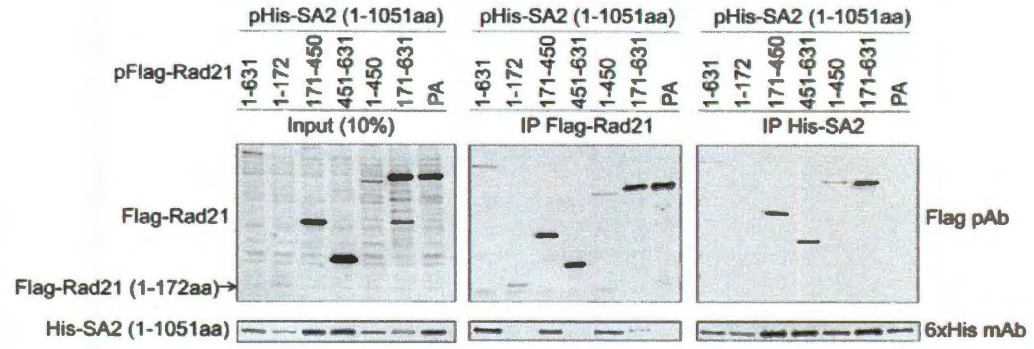


Figure 4.3 Rad21 interacts with SA2 through its middle region (383-392aa). Rad21 interacts with SA2 through the middle region (171-450aa). His-SA2 (1-1051aa) was expressed along with Flag-Rad21 WT and deletion mutants and the complex was co-purified with Ni-NTA or Flag beads. SA2 (1-1051aa) co-expressed with Flag tagged PA protein was used as a negative control.

To further narrow down the polypeptide region of Rad21 that interacts with SA2, we generated baculoviruses overexpressing progressive Rad21 deletion mutants with ~35 amino acid increments/decrements from either the N- or C-terminus of the Rad21 middle region (171-450aa) (Figure 4.4a). SA2 (1-1051aa) along with each of the Rad21 deletion mutants were expressed in Sf21 cells. The complex was then purified using either Ni-NTA or Flag beads and analyzed by western blot. Sf21 cells infected with SA2 (1-1051aa) and Flag tagged PA was used as a negative control. As shown in Figure 4.4b, Rad21

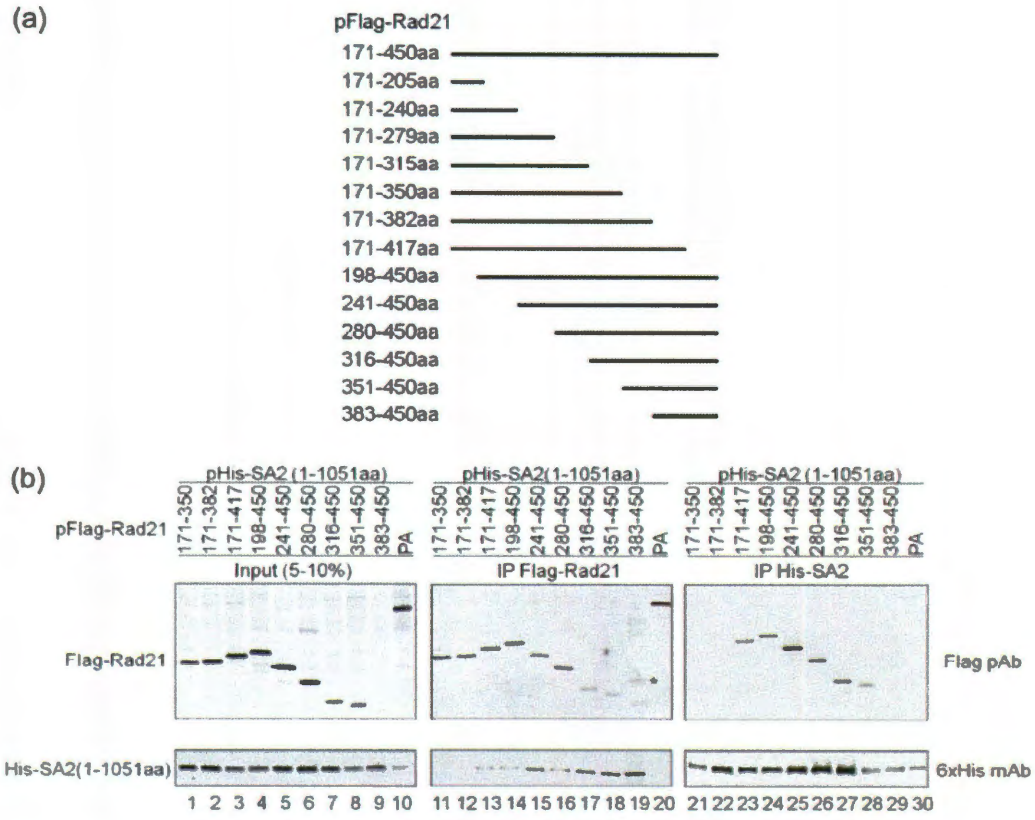


Figure 4.4 Rad21 (171-382aa) does not interact with SA2. (a) Schematic illustration of the middle portion Rad21 deletion constructs made in the baculovirus system. (b) His-SA2 (1-1051aa) was expressed along with the Flag tagged Rad21 deletion mutants, and the complex was co-purified with Ni-NTA or Flag beads. Contaminating antibody bands are marked by asterisks.

(351-450aa), but not Rad21 (171-382aa), forms a complex with SA2 (1-1051aa) (Figure 4.4b, lanes 18 and 28 vs. lanes 12 and 22). Another mutant of interest, Rad21 (383-450aa), was also co-purified with SA2 (1-1051aa) using Flag-beads (Figure 4.4b, lane 19), but the efficiency of SA2 to co-purify Rad21 (383-450aa) using Ni-NTA beads was low (Figure 4.4b, lane 29), likely due to the low expression level of Flag-Rad21 (383-450aa) (Figure 4.4b, lane 9). These results indicated that 351-450aa of Rad21 is sufficient for its interaction with SA2, and that Rad21 (383-450aa) is the minimal interacting region to form a SA2-Rad21 complex *in vitro*.

4.3 Purification and crystallization of the SA1/2:Rad21 complex

To further investigate if the human SA2 binds directly to Rad21 and to verify the interacting domains we have identified in SA2 and Rad21, we co-expressed SA2 (1-1051aa) and Rad21 (171-450aa) and purified the complex from Sf21 cells, in which no other human proteins were present. Approximately 1 x 10^9 Sf21 cells were co-infected with baculoviruses overexpressing 6xHis tagged SA2 (1-1051aa) and Flag tagged Rad21 (171-450aa) at an MOI of 2:4. Sf21 cells were harvested 60h post-infection. Cell pellets were suspended in 60ml of Ni-NTA buffer A. The cell pellets were lysed by sonication and then centrifuged at 35,000g for 40min. The supernatant was then collected and loaded onto 2ml of Ni-NTA resin which was pre-equilibrated with Ni-NTA buffer A. The mixture was gently stirred at 4°C for one hour and then centrifuged at 1,000g for 3min to separate the beads from the flow through. The resin was washed by 25ml of Ni-NTA buffer A. After the wash, the complex was eluted by Ni-NTA buffer B. The eluted

protein solution containing the complex was diluted three times by FPLC buffer A before being loaded onto a 5ml Hitrap Q column. Samples were eluted using an increasing gradient of salt at pH 7.5. The complex was eluted at 400mM NaCl. Fractions containing the complex were collected, concentrated, and loaded onto a 24ml Superose6 gel filtration column. A gel filtration buffer (25mM Tris-HCl pH=7.5, 200mM NaCl, and 5% glycerol) was used to elute the protein. The complex was eluted as a single peak in the gel filtration chromatogram at a position earlier than SA2 (1-1051aa) alone (Figure 4.5). The sample was more than 90% pure. To determine the ratio of the two proteins, the sample was analyzed by Coomassie blue G-250 stained SDS-PAGE. The intensity of the bands representing each of the protein was quantified using ImageJ (Rasband, rsbweb.nih.gov/ij/, 1997-2008) and then normalized by the number of amino acids of each protein. The estimated molar ratio of SA2 (1-1051aa) and Rad21 (171-450aa) was 1:1.

The SA2 (1-1051aa) ortholog, SA1 (1-1055aa) was also co-purified with Rad21 (171-450aa) using the same procedure as described above. Similarly, the SA1 (1-1055aa):Rad21 (171-450aa) complex was eluted as a single peak from a gel filtration column earlier than SA1 (1-1055aa) alone (Figure 4.6). The molar ratio of the two proteins was also estimated to be 1:1. Approximately 1mg SA1 (1-1055aa): Rad21 (171-450aa) could be purified from 1L insect cells at the purity of >90%, which was sufficient for crystallization screenings. Unfortunately, the complex precipitated in more than 80% of the conditions using the Classic Lite kit (Hampton research) even at 2mg/ml. It also precipitated heavily in multiple other screening conditions.

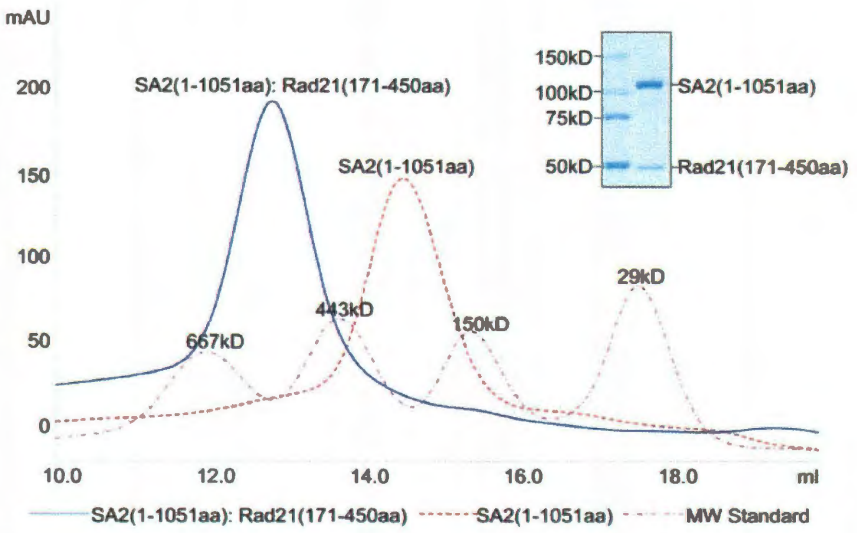


Figure 4.5 SA2 (1-1051aa) and Rad21 (171-450aa) form a stable complex. Gel filtration chromatogram for the SA2 (1-1051aa): Rad21 (171-450aa) complex. Inset shows the Coomassie-stained gel of the SA2:Rad21 complex purified by gel filtration.

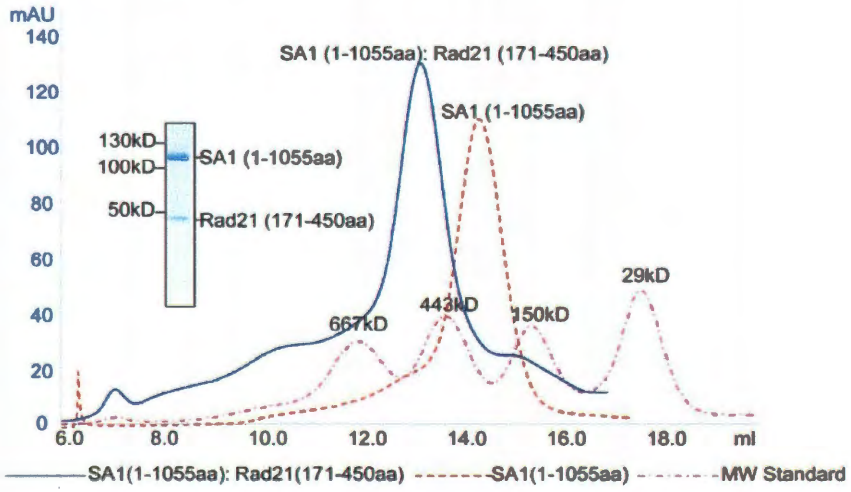


Figure 4.6 SA1 (1-1055aa) and Rad21 (171-450aa) form a stable complex. Gel filtration chromatogram for the SA1 (1-1055aa): Rad21 (171-450aa) complex. Inset shows the Coomassie-stained gel of the SA1:Rad21 complex purified by gel filtration.

Estimating the ratio of SA2 and Rad21 in the complex from a Coomassie stained gel is a rough method, more accurate approaches are required to determine the stoichiometry of complex. We employed different methods, including directly visualizing the complex with TEM, using AU to determine the MW, and purifying the complex containing two differently tagged Rad21.

4.4 EM studies of the SA2:Rad21 complex

Purified SA2 (1-1051aa): Rad21 (171-450aa) was visualized by a transmission electron microscope (TEM). Negative staining method was used to prepare the samples for TEM. The purpose of negative staining is to embed protein molecules in heavy atom stains (e.g., uranium, tungsten, etc.) in order to enhance the contrast between the sample and the background for observation. To prepare a protein sample by negative staining, a negative staining solution (0.75% uranyl formate (Polysciences, Inc.), with the addition of 5M NaOH to reach a dark yellow appearance) was freshly made and filtered with a 0.22 μm filter (Millipore). Copper grids coated with carbon (Polysciences, Inc.) were placed in a vacuum chamber and glow discharged at 5mA for 1min to increase the hydrophilicity of the carbon surface. 3μl of SA2 (1-1051aa):Rad21 (171-450aa) at a concentration of approximately 0.05mg/ml was loaded onto the discharged copper grid. Excess sample was removed after 1min with filter paper by drawing round the edge of the copper grid. The grid was then washed twice with MQ water, once with a drop of staining solution, and stained in a second drop of staining solution for 45s. Excess staining solution was removed with filter paper and the grid was air-dried before use.

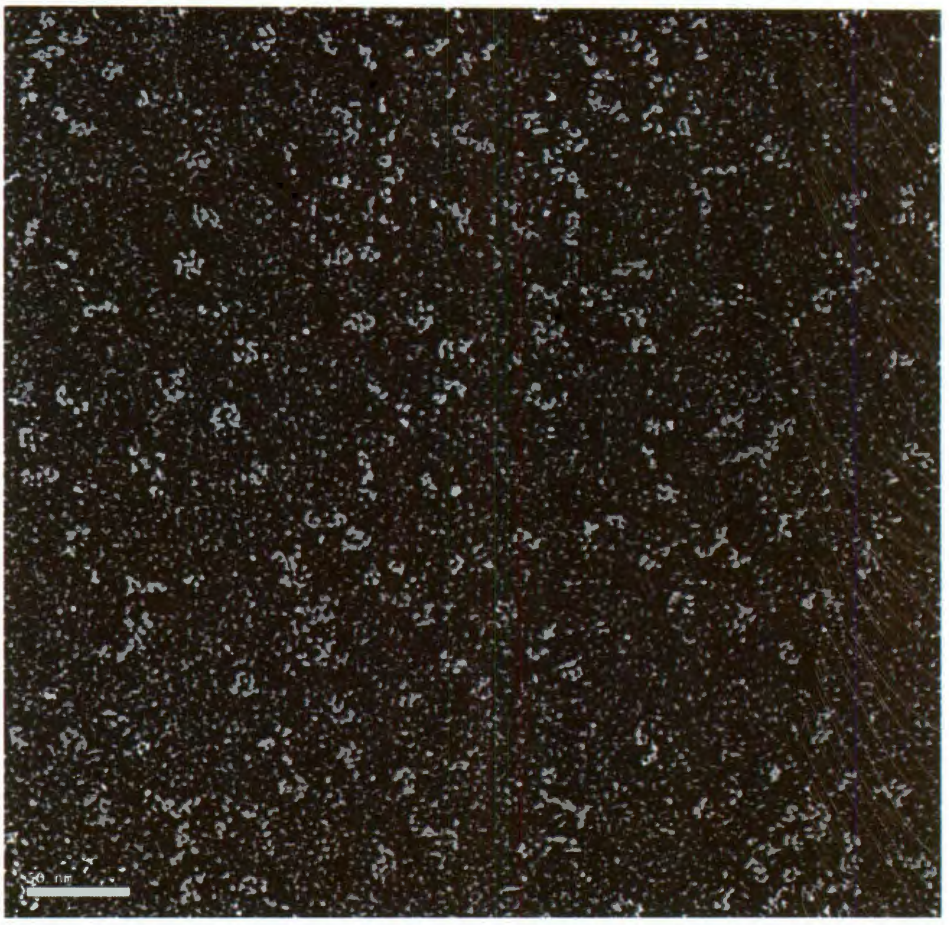
A Joel 2010 TEM was used to image the samples. Micrograph of the SA2 (1-1051aa): Rad21 (171-450) complex sample was captured at a magnification of 50000 X (Figure 4.7a). The magnified view of the particles (Figure 4.7b) revealed that the complex contained multiple domains. The number of domains seemed to vary slightly between the particles, which might be a result of the difference in orientation. Since Rad21 (171-450aa) is a small protein which is predicted to be largely unstructured, it might only occupy one of the domains as shown in the micrograph. The SA2 (1-1051aa) protein contains roughly three more times of amino acids than Rad21 (171-450aa), therefore, it might occupy the rest of the domains. Even there are two molecules of Rad21 (171-450aa) in each complex, SA2 (1-1051aa) is still likely to contain 5~6 domains (Figure 4.7b). To assign the domains to each protein, we can add an anti-6 x His mAb to the complex sample. The 6xHis antibody targets the N-terminal of SA2 (1-1051aa), while SA2 (303-750aa) was found to be interacting with Rad21, therefore the addition of the antibody should not interfere with the interaction between SA2 and Rad21.

We were not able to determine the stoichiometry of the complex from the EM micrograph. It is likely that negative staining method could cause the collapse of the complex structure in the sample preparation. This issue can be addressed using cryo-EM as an alternative method.

4.5 Stoichiometry studies of the SA1/2:Rad21 complex

We have shown that Rad21 and SA1/2 form a stable complex. However, the stoichiometry of the proteins in the complex remains unknown. The one ring model

(a)



(b)

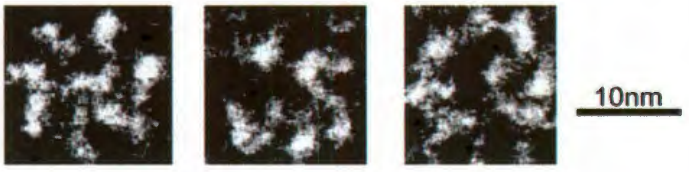


Figure 4.7 TEM picture of the purified SA2 (1-1051aa):Rad21 (171-450aa) complex. (a) A general view of the SA2 (1-1051aa): Rad21 (171-450aa). Scale bar: 50nm. (b) Magnified view of the general shape of the complex. Scale bar: 10nm.

suggests that each cohesin complex contains one molecular copy for each of the four cohesin core subunits. Therefore, SA1/2 and Rad21 should form a complex at the 1:1 ratio. The handcuff model proposes that two Rad21 molecules interact with each other in presence of SA1/2 (Zhang *et al.*, 2008), suggesting that SA1/2 and Rad21 may form a complex at the 1:2 ratio. To differentiate between these two models, I used analytical ultracentrifugation to determine the molecular stoichiometry of the SA1/2: Rad21 complex by measuring the MW of the complex.

SV experiments were performed on both SA2 (1-1051aa) and the SA2 (1-1051aa):Rad21 (171-450aa) complex using a Beckman Coulter Optima XL-A analytical ultracentrifuge. The OD values of the samples were obtained against a water reference at 230nm. The sedimentation velocity experiments were performed at 30,000rpm (AN60 Rotor) and 4°C for 15h. One hundred and fifty scans were recorded for each sample at a radial step size of 30μm. Data analysis was performed as described in Section 2.8. The G(s) of SA2 and the SA2 (1-1051aa): Rad21 (171-450aa) complex are shown in Figure 4.8. The SA2 (1-1051aa):Rad21 (171-450aa) complex showed larger value of sedimentation coefficient as compared to SA2 (1-1051aa) alone, indicating that SA2 (1-1051aa) and Rad21 (171-450aa) indeed form a stable complex.

To determine the MW of the SA2 (1-1051aa):Rad21 (171-450aa) complex, the same approach was used as described above. The G(s) of the complex at the two different concentrations are shown in Figure 4.9a, suggesting no oligomerization during data collection. The global Monte-Carlo analysis showed that there was one major species with the molecular weight of 151.8 (144.2, 168.1) and a frictional ratio of 1.82 (1.72, 2.22) (in parenthesis: 95% confidence intervals). This species should be the complex. If

SA2 and Rad21 form a complex at 1:1 ratio, the theoretical MW is 154.6. If the complex is formed at 1:2 ratio, the theoretical MW is 187.4. Based on the calculated MW, Rad21 and SA2 should form a 1:1 complex. However, if the complex contained a trace amount of 1:2 complex, the AU method might not be accurate enough to detect it. To address this question, 6 x His tagged Rad21 and Flag tagged Rad21 were co-expressed with HA tagged SA2 and purified in the order of Ni-NTA column, anion exchange column and gel filtration column. If SA2 and Rad21 do form a 1:2 complex, Flag tagged Rad21 will be detected in the fractions collected from the major peak of the gel filtration. The fractions from the gel filtration column were analyzed by anti Flag mAb. We did observe a small amount of Flag tagged Rad21 (171-450aa) in the fractions from the major peak (Figure 4.10a, lanes 3-6) but the amount of the Flag tagged Rad21 (171-450aa) in the aggregation peak (Figure 4.10a, lane 1) was much more significant consistent with the observation in section 3.3 that Rad21 (171-450aa) had a strong tendency to form aggregates. The small amount of Flag tagged Rad21 (171-450aa) found in fractions around the major peak may reflect the existence of a small amount of the 1:2 SA2: Rad21 complex. It was suggested that *in vivo* the population of the dimerized cohesin rings is much smaller comparing to the population of the single cohesin ring (Zhang *et al.*, 2008). On the other hand, it is also possible that the small amount of Flag tagged Rad21 (171-450aa) detected in the major peak might be due to leaking of aggregated Rad21 (171-450aa) from the aggregation peak (Figure 4.10b).

We have also performed sedimentation velocity experiment for the SA1 (1-1055aa):Rad21 (171-450aa) complex. The G(s) of the complex at the two different concentrations are shown in Figure 4.9b, suggesting no oligomerization during data

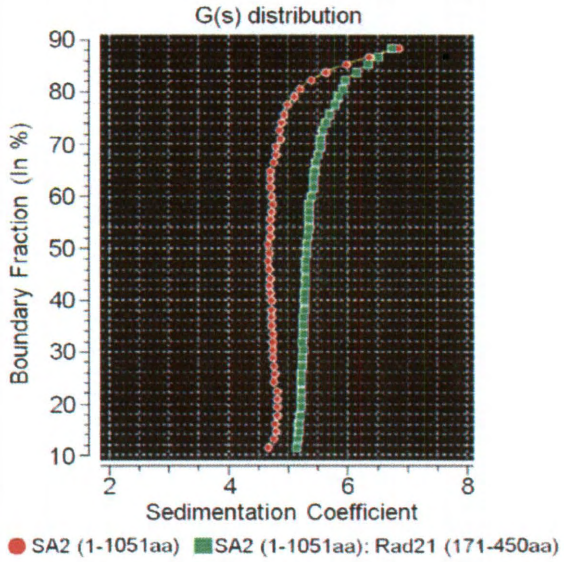


Figure 4.8 Velocity sedimentation results for SA2 (1-1051aa) and the SA2 (1-1051aa): Rad21 (171-450aa) complex. The complex shows an increase in the sedimentation coefficient compared to SA2 (1-1051aa) alone.

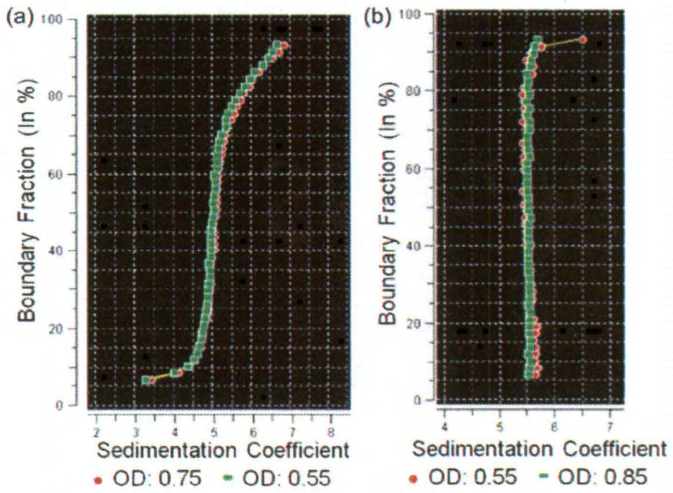


Figure 4.9 G(s) for different samples. (a) G(s) for SA2 (1-1051aa):Rad21 (171-450aa). (b) G (s) SA1 (1-1055aa):Rad21 (171-450aa).

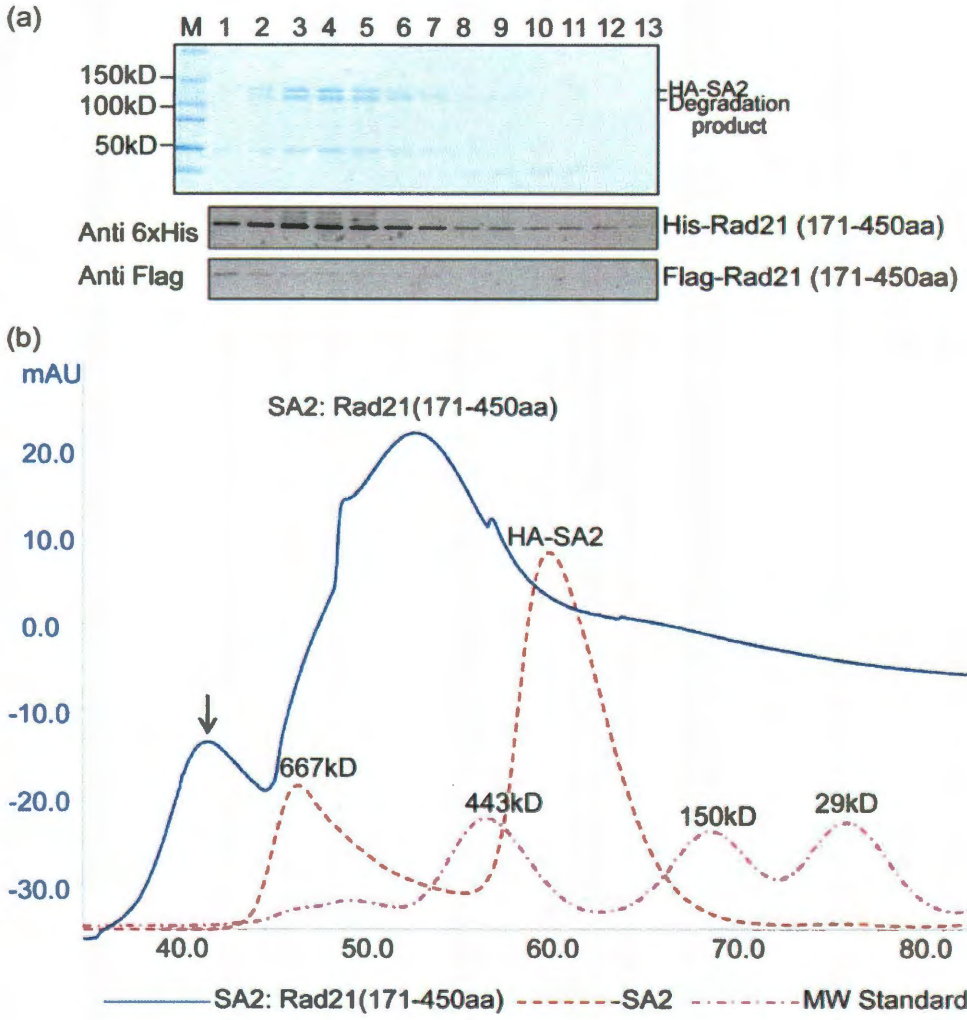


Figure 4.10 Co-purification of the two differently tagged Rad21 with HA tagged SA2. (a) Fractions from the gel filtration chromatogram were analyzed using Coomassie staining (upper panel), anti 6 x His mAb (middle panel), and anti Flag mAb (lower panel). (b) Gel filtration chromatogram of the SA2:Rad21 complex. Lanes 1from (a) are collected at the position indicated by the arrow and lanes 3-6 are collected at the elution peak for the SA2: Rad21 (171-450aa) complex.

collection The global Monte-Carlo analysis showed that there was one major species with the molecular weight of 166.2 (164.2, 168.7) kD and a frictional ratio of 1.97 (1.96, 2.00). If SA1 (1-1055aa) and Rad21 (171-450aa) formed the complex at 1:1 ratio, the calculated MW would be 155.0 kD. For a 1:2 complex, the calculated MW would be 187.8. The predicted the MW was closer to 155.0kDa. Therefore, like SA2, SA1 should form a complex with Rad21 (171-450aa) at 1:1 ratio.

Our *in vitro* experiment showed that Rad21 and SA2 form a 1:1 complex instead of 2:1 as suggested by the handcuff model. There are a few explanations for this contradiction. First of all, it is possible that the dimerization of Rad21 requires other human proteins in addition to SA2. The two molecules of Rad21 may not directly interact, but SA2 might help to recruit other cohesin associated proteins to facilitate the dimerization of Rad21. Those associated proteins would be missing in the insect cell system. Therefore, SA2 and Rad21 only form a complex at 1:1 ratio when co-expressed in Sf21 cells. The second possibility is that the integrity of the Rad21 structure is necessary for the formation of a 2:1 complex. Due to the low expression level of the full length Rad21, only the Rad21 (171-450aa) was tested for the complex formation. Although the N- and C-terminus of Rad21 do not directly interact with SA2, they might still play a part in the dimerization. Because it is not known if SA2 is the only protein that directly triggers the dimerization of Rad21, it is possible that the dimerization requires SA2 to recruit other proteins, which triggers the dimerization through associating with Rad21 through the N- or C- terminus of Rad21. Without N- or C-terminus, the association is not able to occur and thus, there is no dimerization. The handcuff model suggested that the two Rad21 molecules interact with each other in an anti-parallel

manner (Zhang *et al.*, 2008). Therefore, it is also possible that the interaction between Rad21 N- and C -termini contributes to part of the dimerization, without which, the dimerization is prohibited.

Chapter V. Characterization of the interaction between SA2 and Rad21 in mammalian cells

In the previous chapter, I have identified that SA2 (1-1051aa) is sufficient for interacting with Rad21 and Rad21 (383-450aa) is the minimal region identified for interacting with SA2. To further narrow down the region of Rad21 that interacts with SA2 and identify the key amino acids within the region will provide a better understanding of the Rad21-SA2 interaction.

5.1 Define the amino acid residues of Rad21 essential for interaction with SA2

To further narrow down the region of Rad21 that interacts with SA2, additional experiments were performed in a mammalian cell expression system. The baculovirus system used in previous chapters is a more convenient and lower costly system when it comes to large scale protein expression and purification. However, the experimental cycle from cloning the genes of interest to expressing the protein is much longer in insect cells compared to that in mammalian cells (a week vs. a month). It takes a few rounds of transfection and co-immunoprecipitation (IP) experiments to narrow down the region of interest and identify the key amino acids in Rad21 for interaction with SA2. Therefore, a mammalian cell expression system was chosen as a more efficient system for the experiments in this chapter. In addition, the cell line used in this chapter is the Human embryonic kidney (HEK) 293T cells. After the essential amino acids of Rad21 are identified in co-IP assays, the results will also be confirmed through IP of the endogenous

cohesin components. Therefore, 293T, as a human cell line, is the ideal cell line for the experiments to be carried out in this chapter.

SA2 (1-1051aa) was cloned into the pCS2 MT vector bearing a 6xMyc tag at the N-terminus. A series of Rad21 deletions mutants were designed based upon the results from section 4.2 and were cloned into the pFlag CMV2 vector (Figure 5.1a). The DNA used for each transfection was prepared by midi-prep from 100ml DH5α cells using the Qiagen Midi-kit. The minimal concentration of DNA used for transfection was 400ng/μl. SA2 (1-1051aa) was transfected along with each of the Rad21 deletion mutants into 293T cells. To perform the transfection, each 10cm tissue culture plate was seeded with 5 x 10^6 293T cells in DMEM medium supplemented with 10% FBS. Approximately 5μg of DNA was mixed with 50ul 2.5M $CaCl_2$ solution, 450ul MQ water and 500ul 2 x Hank's buffered salt solution (HBSS) before being incubated at room temperature for 20 minutes. The mixture was then added to each plate and the cells were incubated at 37°C. Medium was changed 16h post transfection and the cells were harvested 40h post-transfection. Cells transfected with SA2 (1-1051aa) along with the pFlag CMV2 empty vector (EV) was used as a negative control.

To harvest the cells, medium was discarded and the cells were washed twice with PBS buffer. 1ml lysis buffer (50mM Tris-HCl, pH=7.4, 1% Triton X-100, 150mM NaCl, 1mM EDTA, Complete mini Protease Inhibitor cocktail tablet (Roche; 1 tablet in every 10ml lysis buffer)) was then added to the cells. With the addition of the lysis buffer, the cells were detached from the bottom of the plates and transferred to 1.5ml eppendorf tubes. The cells were pipetted up and down for 20 times to facilitate the breaking of the nuclear membranes, followed by on-ice incubation for 10min. The cells were then spun

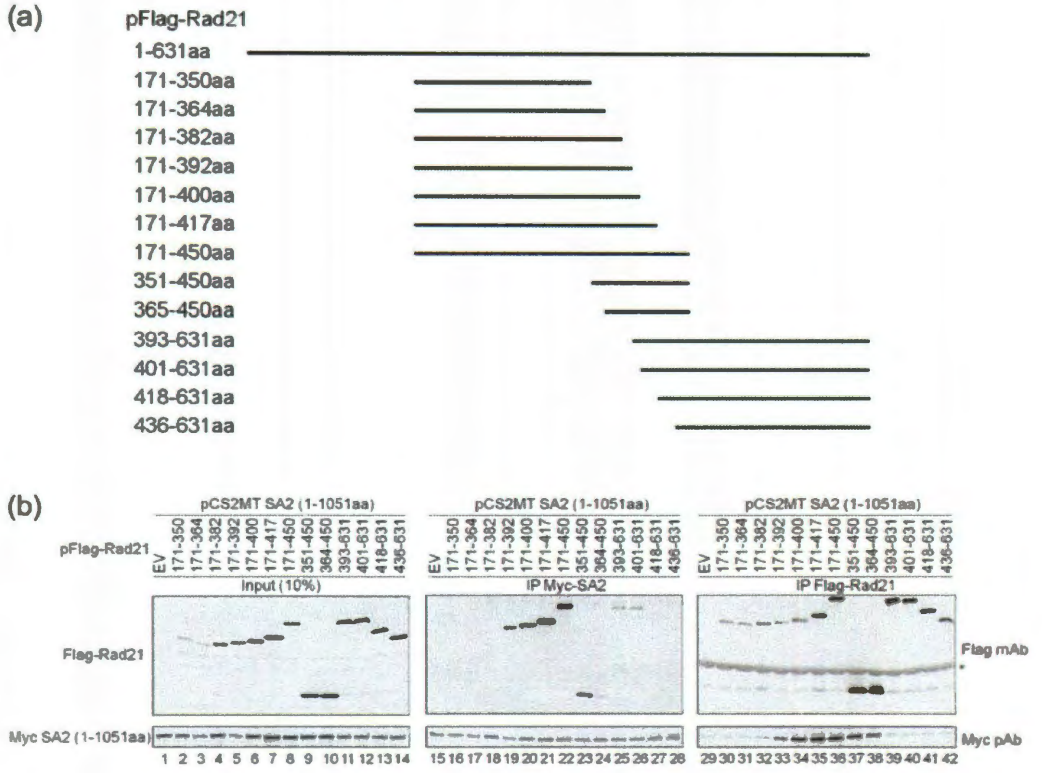


Figure 5.1 Rad21 interacts with SA2 through a 10aa region. (a) Schematic illustrations of Rad21 deletion constructs in the context of the full length Rad21 in the mammalian expression vector pFlagCMV2. (b) The Rad21 383-392aa region is critical for interacting with SA2. Myc-SA2 (1-1051aa) was co-transfected along with the Flag-Rad21 deletion mutants and immunoprecipitated with Flag or Myc beads and probed with either the Myc polyclonal antibody (Myc pAb) or the FLAG mAb. Flag empty vector (EV) was used as a negative control.

down at 14,000g for 15min. The supernatants were collected and the concentration of each sample was measured using Bradford assay. Approximately 1.2mg protein from each transfection was added to either Flag beads or Myc mAb conjugated agarose bead (Myc beads; Sigma-Aldrich). To perform the Co-IP, samples were incubated with either Flag or Myc beads with constant mixing at 4°C for 3h. The beads were then washed three times with TBS buffer and the tagged proteins were eluted with 2 x SDS sample buffer.

The Co-IP results were analyzed using western blot (Figure 5.1b). Rad21 (171-392aa) was able to pull down SA2 (1-1051aa) while Rad21 (171-382aa) and Rad21 (393-631aa) were not (Figure 5.1b, Lanes 19 and 33 vs. lanes 18 & 32 and lanes 25 &39). This is consistent with the results obtained from the insect cells. The dramatic differences in the binding affinity of Rad21 (171-382aa) and Rad21 (171-392aa) with SA2 also lead to the conclusion that Rad21 383-392aa is crucial for interacting with SA2. A secondary structure prediction was performed for the Rad21 protein as mentioned in section 3.5. The structure prediction showed that the Rad21 middle region (171-450aa) was largely unstructured with three predicted α-helices, and one of them is predicted with very low probability (Figure 5.2a). Interestingly, one of the two α-helices with high probability was formed by 383-392aa. A deletion mutant devoid of the 10 amino acids from the full length Rad21 was cloned into the pFlag CMV2 vector. An alignment of Rad21 protein in different species have shown that in the 383-392aa region, amino acids L^{385}, F^{389} and T^{390} are conserved (Figure 5.2b). Therefore, site mutations were designed according to the conserved amino acids and cloned into the pFlag CMV2 vector.

SA2 (1-1051aa) was transfected into 293T cells along with each of the Rad21 mutants and Co-IP was performed with both Flag and Myc beads as described above and

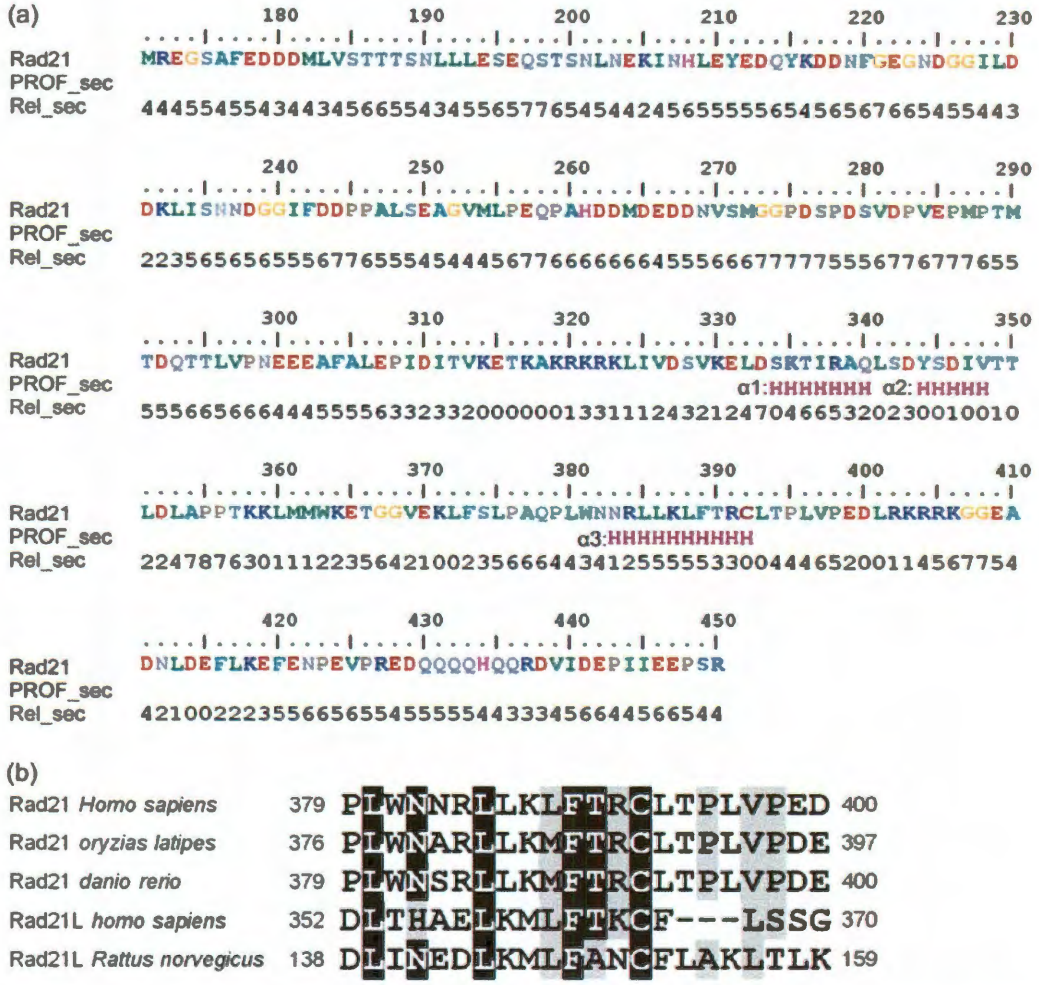


Figure 5.2 Rad21 383-392aa form an α-helix and amino acids L^{385}, F^{389} and T^{390} are conserved. (a) Rad21 amino acid sequence is colored based on residue types (e.g. blue for positively charged, red for negatively charged, green for hydrophobic, silver for polar, etc). PROF_sec predicts the secondary structure (H= Helix). Rel_Sec shows the reliability index of the PROF_sec prediction (0=low, 9=high). (b) Sequence alignment of Rad21 from various vertebrate species. The sequence alignment was prepared using BioEdit (Hall, 1999). The conserved Rad21 L^{385}, F^{389} and T^{390} residues were used for making site-directed mutations.

analyzed by western blot. Cells transfected with SA2 (1-1051aa) and the Flag EV was used as a negative control. The western blot showed that the mutation of L^{385} and F^{389} significantly reduced the binding of Rad21 to SA2 (Figure 5.3, lanes 18 and 30). Furthermore, the simultaneous mutation of all the three amino acids, L385, F389 and T390 or the deletion of 383-392aa completely abrogated the Rad21-SA2 interaction. (Figure 5.3, Lanes 21 & 33 and 22 & 34).

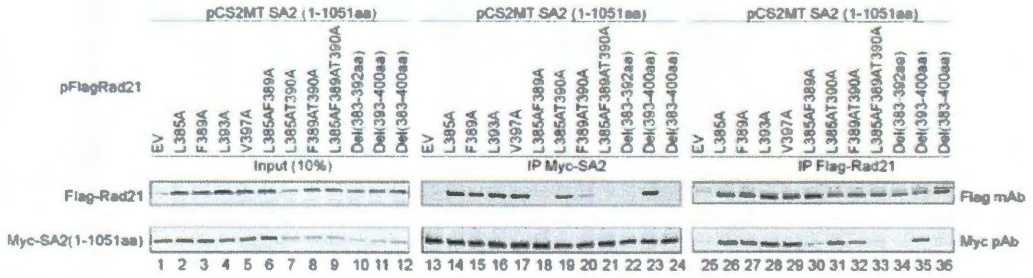


Figure 5.3 L^{385} and F^{389} are critical for Rad21 to interact with SA2 . Rad21 L^{385}, F^{389}, and T^{390} are important for interaction with SA2.

To confirm if the mutations identified above are able to disrupt the Rad21-SA2 interaction, Myc-Rad21 wildtype (WT) or mutant Rad21 were cloned into the pCS2 MT vector and ectopically expressed in 293T cells. Immunoprecipitation of SA2 by WT or mutant Rad21 was analyzed by western blot. The results showed that the immunoprecipitation of endogenous SA2 was significantly reduced in Rad21 mutants (Figure 5.4a), which is in consistence with the above finding that amino acids L^{385}, F^{389} and T^{390} are essential for the Rad21-SA2 interaction.

The SA2 ortholog, SA1, shares 70% sequence identity with SA2. A previous study has showed that the Rad21 (362-404aa) is responsible for interacting with SA1 (Shintomi *et al.*, 2009) *in vitro*. The Rad21 (383-392aa) locates within the 362-404 region. Therefore, it was interesting to examine whether the Rad21 (383-392aa) and its specific binding sites are also responsible for interacting with SA2. Myc tagged SA1 (courtsey of Dr. Debananda Pati) was transfected along with Flag tagged wildtype or mutant Rad21 into 293T cells and Co-IP was performed using Flag beads as described above. Cells transfected with Myc tagged SA1 and the pFlagCMV2 EV was used as a negative control. Similar to SA2, mutation of L^{385} and F^{389} significantly reduced the binding of Rad21 to SA2 (Figure 5.4b). Furthermore, the simultaneous mutation of all the three amino acids, L^{385}, F^{389} and T^{390} or the deletion of 383-392aa completely abrogated the Rad21-SA2 interaction (Figure 5.4b). To confirm this result with the endogenous SA1, the IP products pulled down by Myc-tagged Rad21 proteins from the above experiments were analyzed by Western blot. As expected, the immunoprecipitation of SA1 was reduced significantly in the Rad21 mutants compared to Rad21 WT control (Figure 5.4a). Therefore, the Rad21 (383-392aa) is not only responsible for interacting with SA2, but also play an essential in interacting with SA1.

To examine if the disruption of SA2 binding to Rad21 has any effect on Rad21 association with the other cohesin core subunits, the immunoprecipitation of both Smc1 and Smc3 were also analyzed by western blot. Unlike SA1/2, both Smc1 and Smc3 remain unchanged in all Rad21 mutants comparing to the WT control (Figure 5.4a), indicating that the disruption of Rad21-SA2 binding has no apparent effect on Smc1-Smc3-Rad21 interaction.

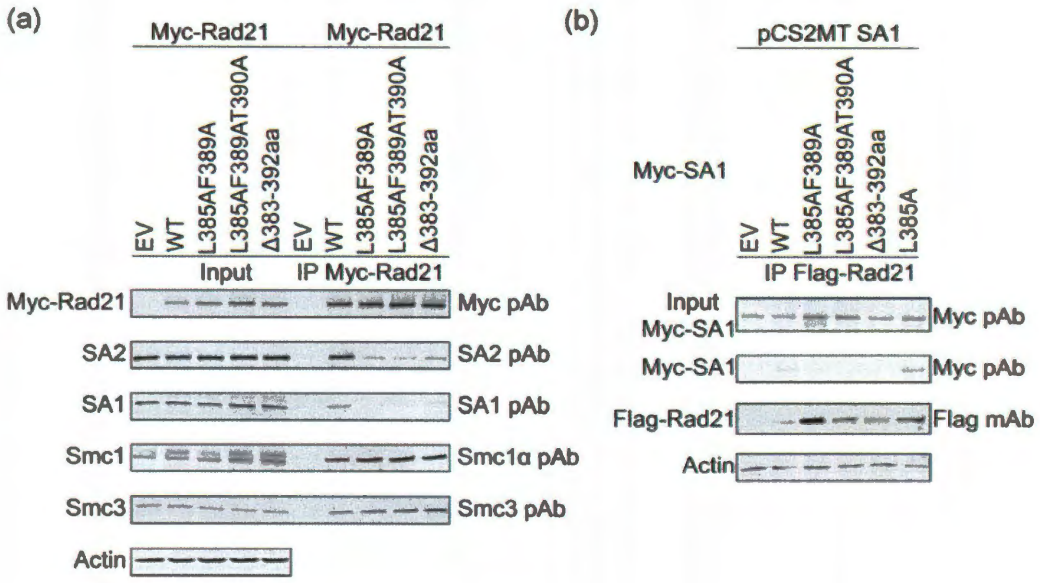


Figure 5.4 The association of the Rad21 mutants and the cohesin subunits. (a) The amount of SA1/SA2 immunoprecipitated by the Rad21 mutants was significantly decreased while the amount of Smc1 and Smc3 remained constant. (b) The Rad21 383-392aa region is critical for interacting with SA1 *in vitro*. Myc-SA1 was co-transfected along with the Flag-Rad21 deletion mutants and immunoprecipitated with Flag beads and probed with either the Myc polyclonal antibody (Myc pAb) or the FLAG mAb. Flag empty vector (EV) was used as a negative control.

5.2 Disruption of the Rad21-SA2 interaction leads to premature separation of sister chromatids

To examine the physiological consequence of the disruption of the Rad21-SA2 interaction, Myc-tagged WT and mutant Rad21 were expressed in 293T cells. Myc EV was used as a control. To examine the effect of the ectopic protein, endogenous Rad21 was knocked down using Rad21 3'-UTR siRNA (Qiagen) 24h before transfection. Silencer Negative Control siRNA (Applied Biosystems) was used as a negative control. To enhance the efficiency of the knockdown, cells were treated with siRNA again 24h after transfection. Cells were harvest 72h after the second siRNA treatment. Three hours prior to the harvest, colcine was added to arrest the cells in metaphase by inhibiting microtubule polymerization through binding to tubulin. To harvest the cells, cells were treated with trypsin-EDTA solution (0.25% (w/v) trypsin and 0.2% (w/v) EDTA prepared in 1 x HBSS solution) for 5min. Trypsin digests the adhesion proteins in cell-cell and cell-matrix interactions, while EDTA, as a calcium chelator, which integrins needs to interact with other proteins for cell adhesion and also 'wipes' out the calcium in DMEM which functions as a trypsin inhibitor. Therefore, the cells are dettached after trypsin-EDTA treatment. The cells were then centrifuged at 800g for 6min. The expression of WT and mutant Rad21 as well as endogenous Rad21 following siRNA treatment are shown in Figure 5.5a.

To perform metaphase chromosome spread analysis, 10ml of pre-warmed hypotonic solution (0.075M KCl) was added to gently agitate the pellets, followed by 15 min incubation at 37°C to break the membrane. 1ml of fixative (methanol: glacier acetic acid= 3:1) was added after incubation and cells were then centrifuged at 800g for 6 min

and supernatants were discarded. 4ml of fixative was added to each sample and incubated at room temperature for 30 min. Fixative was changed twice and 200μl of fixative was added after the final centrifugation and cells were resuspended. To create slides, 30μl was taken from each sample and dropped on angled slides from 10 inches above. The slides were then air-dried and stained in Giemsa solution for 10 min and then rinsed with water ten times. About 100 metaphase cells for each treatment were counted using a Zeiss AxioSkop 40 microscope (Carl Zeiss).

The cells were divided into four categories: 1) chromatids with arm cohesion; 2) chromatids without arm cohesion but still linked at the centromeres; 3) chromatids separated but still in alignment with each other; and 4) single chromatids, randomly scattered (Figure 5.5b). In wildtype cells, around 10% of the cells showed cohesion defects. In the siRNA treated cells, approximately 50% precocious sister chromatid separation was noted in the Myc-EV control treated with Rad21 3'-UTR siRNA, while 33% of cohesion defects were observed in cells with the ectopic expression of Rad21. Therefore, the Rad21 depletion phenotype could be partially rescued by the overexpression of the WT Rad21 (Figure 5.5c). However, in the same assay, Rad21 $L^{385A}F^{389A}$ mutants had only about 5% rescue rate. Instead of rescuing, Rad21 Δ (383-392aa) caused more precocious sister chromatid separation than the Myc-EV, possibly due to a dominant negative effect of the deletion mutant. The dominant negative effect of Rad21 Δ (383-392aa) was also apparent from the cells transfected by WT and Rad21Δ (383-392aa), which showed a rescue rate slightly lower than that of the cells treated by WT Rad21 only (Figure 5.5c). These results confirmed that amino acids 383-392 in

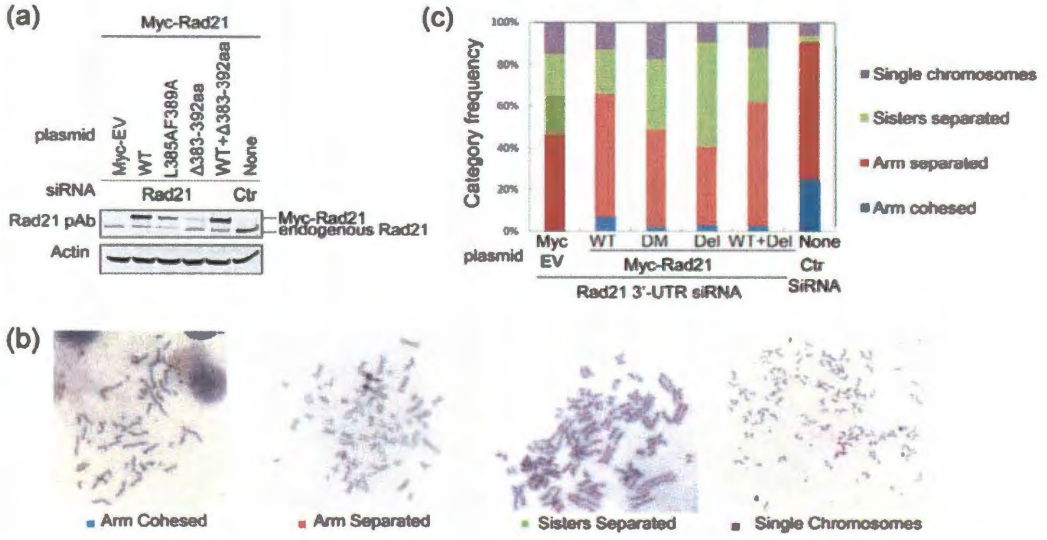


Figure 5.5 Disruption of Rad21 binding to SA2 results in premature separation of sister chromatids. (a) Expression of Myc-Rad21 WT and deletion mutants in 293T cells. 293T cells were treated with Rad21 3'-UTR or a control (Ctr) siRNA 24h before and after being transfected with Rad21 WT or mutant plasmids. The expression level of Myc-Rad21 was analyzed by Western blot using Rad21 polyclonal antibody (Rad21 pAb). (b-c) Rad21 mutations cause premature separation of sister chromatids. For metaphase chromosome spread analysis, ~100 mitotic cells were counted for each treatment according to the categories shown in (b). The frequency of mitotic cells was calculated and plotted in (c). DM, double mutant $L^{385A}F^{389A}$; Del, Δ383-392; WT (Wild Type) + Del, WT + Δ383-392.

Rad21 are not only required for the Rad21-SA2 interaction but also physiologically important for sister chromatid cohesion.

5.3 Model on the role of the Rad21-SA2 interaction in cohesin ring maintenance

Rad21 (383-392aa) has been identified to be important for maintaining the normal chromatid activity. However, it is not clear how the interruption of the Rad21-SA2 interaction leads to the early separation of sister chromatids.

According to the handcuff model, two Rad21 molecules interact with each other in the presence of the SA1/2. Therefore, the interruption of the Rad21-SA1/2 interaction may prevent the formation of the cohesin ring dimerization. We analyzed the IP products by Myc-tagged Rad21 from section 5.1 by western blot to determine if Rad21 molecules still pull down when the interaction with SA2 is disrupted. Western blot showed that the Myc-tagged Rad21 was able to pull down endogenous Rad21 (Figure 5.6).

The result has raised two questions. First, according to the handcuff model, the Rad21 molecules interact with each other only in the presence of SA1/SA2 (Zhang *et al.*, 2008). The mutant Rad21 fails to recruit SA1/2 to the cohesin complex, however, the interaction between two mutant Rad21 molecules were not affected. This seems to be a contradictory compared to previous findings. To address this, we noticed that the notion that Rad21 interacts with each other is based upon the observation from the cells in which SA1/2 were knocked down. In our case, SA1/2 still exist in the cells, only failing to bind to the SA1/2 protein. This might indicate that the dimerization of Rad21 molecules is an indirect process. It depends on SA1/2 to recruit other proteins. It is possible that SA1/2

are still localized to the chromatin in the loci that are close enough to the cohesin ring. Therefore, it can still recruit the protein responsible for the Rad21 dimerization. However, this protein remains unknown.

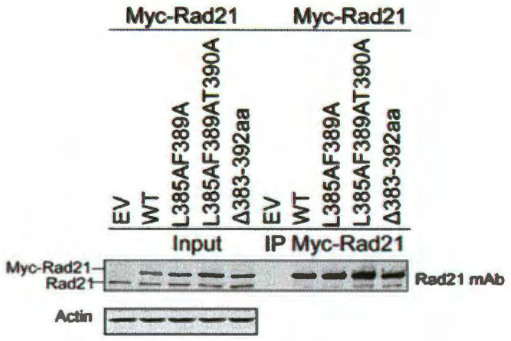


Figure 5.6 The association of Rad21 mutants and the endogenous Rad21. The amount of endogenous Rad21 immunoprecipitated by the Rad21 mutants remained constant.

Second, if the cohesin rings are still associated with each other in the case of mutant Rad21, how does it lead to the early separation of sister chromatids? An explanation is that the early separation of the sister chromatids is not due to the separation of the cohesin rings, but is due to the open of cohesin rings. It was mentioned in section 5.1, that the Rad21-Smc1-Smc3 complex remains intact in the case of mutant Rad21. How is this related to the open ring? Though it has been shown that Smc1 and Smc3 were associated with Rad21 in the absence of SA2 (Haering *et al.*, 2002), there is no conclusive evidence showing that they form a cohesive ring. It is also likely that they do not form a contiguous cohesin ring in the absence of Rad21-SA2 binding. Smc1 and Smc3 form a hinge domain through the center region of each molecule and the head

region, consisting of the N- and C-terminus of each molecule, form an ATPase domain that interacts with the N- and C- terminus of Rad21 respectively (Anderson *et al.*, 2002; Haering *et al.*, 2002). If the hinge domain or one of the ATPase head dissociates, the three subunits will remain connected (Figure 6.1). SA2 is a core subunit identified in the cohesin complex. However, it is not clear what role SA2 plays in holding sister chromatids together. SA2 is a large, 1231aa protein. Our analytical ultracentrifugation data show that SA2 (1-1051aa) has a frictional ratio of 1.66, indicating SA2 is an elongated molecule. Although SA2 does not directly interact with Smc1 or Smc3, we

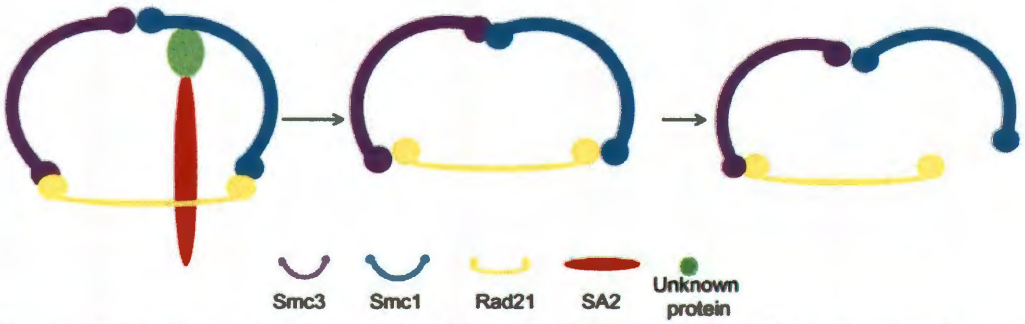


Figure 5.7 Model showing the Rad21-SA2 interaction and SA2 fortifying the cohesin ring. SA2 provides the tension to stabilize the tripartite ring either by itself or through another protein to stabilize the ring structure. Disruption of SA2 binding to Rad21 may lead to the collapse and early dissolution of the ring causing premature separation of sister chromatids.

speculate that SA2 can recruit other cohesin associated protein/s to provide the inner tension required to stabilize the ring (Figure 5.7). When the interaction between SA2 and Rad21 is interrupted by Rad21 mutations, SA2 fails to associate with the cohesin ring, and may be due to its inability to recruit the associated protein/s, leading to the collapse

of the cohesin ring. The collapse of the cohesin ring would fail to generate sister chromatid cohesion and result in the premature separation of sister chromatids. Since SA2 interacts with Rad21 at 383-392aa which is close to the C-terminus of Rad21, interruption of the C-terminal Rad21 and Smc1 interaction in mutant Rad21 is also possible. How SA2 strengthens and stabilizes the cohesin ring is an important question and our present results provide a framework for future studies.

Chapter VI. Conclusions and Perspectives

The cohesin complex has been a subject of intensive study since its core subunits were characterized more than a decade ago (Guacci *et al.*, 1997). Active research has been conducted on the characterization and physiological function studies of the cohesin core subunits and other cohesin related components. However, little is known about the cohesin core protein SA1/SA2 other than its supportive role in the cohesin ring. Our characterization of the Rad21 and SA2 interaction has provided a deeper insight into how physically and physiologically SA2 interacts with the cohesin complex. Our studies provide a model for how SA2 structurally strengthens the cohesin ring through its interaction with Rad21. The conclusions of our studies and future research directions are discussed below.

6.1 Summary of major findings

We have mapped the amino acid domains of the Rad21-SA2 interaction using an array of biochemical and cell biology methods. We found that SA2 physically interacts with Rad21 at multiple polypeptide regions while Rad21 interact with SA2 only through a 10 amino acid conserved region. Deletion of this 10aa α-helix or mutation of specific amino acids within it can interfere with the interaction of Rad21 with SA2 both *in vitro* and *ex vivo*. Importantly, disruption of the Rad21-SA2 interaction by mutated Rad21 proteins leads to premature sister chromatid separation, indicating the structural and physiological importance of this 10aa long helix. We have also performed structural characterization for Rad21 and SA2, the results from which provided directions for

further investigation of the structural basis of protein-protein interaction in the cohesin complex.

6.2 The Rad21-SA2 interaction in human cohesin

It was previously reported that in yeast, SA2 associates with the tripartite cohesin ring through binding to Rad21 (Haering *et al.*, 2002). However, the human Rad21-SA2 interaction at the molecular level and its consequence in chromosomal segregation are lacking. We have used biochemical, biophysical and cell biology methods to characterize the Rad21-SA2 interaction in human cohesin complex.

6.2.1 Details of the Rad21-SA2 interaction

We have shown that SA2 interacts with Rad21 at multiple amino acid domains. In particular, the N-terminal (1-450aa) and the middle portion (581-1051aa) of SA2, two non-overlapping polypeptide regions, are both capable of forming strong interactions with Rad21, indicating that SA2 interacts with Rad21 through a few peptide stretches, and the structural integrity of the whole protein may be important for their interaction. The C-terminal 180aa of SA2 (residues 1052-1231aa) are dispensable for the Rad21-SA2 interaction.

While SA2 interacts with Rad21 through multiple regions, the Rad21 interaction with SA2 is restricted to a 10aa α-helical region. In this 10aa region, L^{385} and F^{389} are next to each other in the helix (Figure 6.1), suggesting that this particular side of the α-

helix interacts with SA2. Considering the localized nature of these important Rad21 residues, it is possible that the multiple regions of SA2 that are implicated in the Rad21-SA2 interaction, although not adjacent in the sequence, may form a contiguous surface for interacting with Rad21.

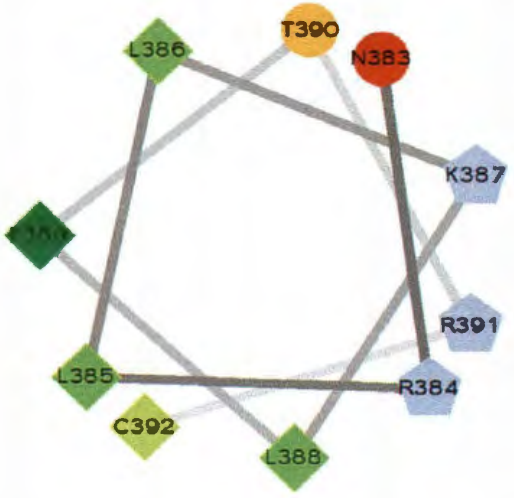


Figure 6.1 Helical wheel illustration of the Rad21 383-392aa. L^{385} and F^{389} are next to each other in the helix (Zidovetzki *et al.*, 2003).

6.2.2 Stoichiometry of the Rad21-SA2 interaction

The handcuff model suggested that two Rad21 molecules interact with each other in the presence of SA1/SA2. Therefore, Rad21 and SA2 may form a 2:1 complex. However, the *in vitro* experiment showed that Rad21 and SA2 form a 1:1 complex instead of 2:1. We have provided a few explanations for this contradiction. First, it is possible that the dimerization of Rad21 does not solely depend on SA2. SA2 might play the role of recruiting other cohesin associated proteins to trigger the dimerization of the

Rad21. Those associated proteins are missing in the insect cell system. Therefore, SA2 and Rad21 only form a complex at 1:1 ratio when co-expressed in the insect cell system. In addition, the integrity of the Rad21 structure is necessary for the formation of a 2:1 complex. Due to the low expression level of the full length Rad21, only the Rad21 (171-450aa) was test for the complex formation. Although the N- and C-terminus of Rad21 do not directly interact with SA2, they might still play a part in the dimerization by interacting with each other or providing a binding region for the protein that directly triggers the dimerization.

6.2.3 Comparison of the Rad21-SA2 interaction between yeast and human

In yeast, Scc3 binds to Scc1 through its C-terminal Separase cleavage fragment (Haering *et al.*, 2002). Hence, it is likely that Scc3 still associates with the cohesin complex after sister chromatid separation. Unlike in yeast, human Rad21 interacts with SA2 through a central 10 amino acid α-helical region which does not overlap with either the N- or C-terminal Separase cleavage products. However, sequence alignment between yeast Scc1 and human Rad21 using BLASTP (www. ncbi.nlm.nih.gov) shows human Rad21 (171-450aa) and the C-terminal cleavage fragment of yeast (269-566aa) have an overlapping region of ~170aa (Figure 3.11), which includes the 10aa region. Therefore, after Rad21 is cleaved by Separase, it is likely that SA2 dissociates from the cohesin complex along with the Rad21 central cleavage product.

It is not known why the yeast and human Scc3 subunits are differently associated with the cohesin complex after Separase cleavage. It may relate to the recycle of the

cohesin subunits. In both yeast and human, the N- and C-termini of Rad21 should dissociate from the Smc1 and Smc3 heterodimer so that the heterodimer can be used as the 'building block' for the re-assembly of the cohesin complex. SA2 and Scc3 are also required to be released by dissociating from the Rad21 binding region to be recycled. In yeast, the dissociation of Scc3 from Rad21 might take place after the dissociation of the C-terminal of Rad21 from Smc1. While in human, the dissociation of SA2 can take place immediately after the cleavage of Rad21, which may allow the free SA2 to perform other functions despite being a subunit of the cohesin complex.

6.3 Characterization of the human SA2 protein

6.3.1 The role SA2 plays in the cohesin model

Earlier studies have shown that Smc1, Smc3 and Rad21 form a tripartite ring while SA2 binds to the ring through Rad21 and the formation of a Smc1-Smc3-Rad21 complex does not require SA2 (Haering *et al.*, 2002). However, it is not clear if SA2 is required for the maintenance of the ring. In our study, we have shown that SA2 is indispensable for maintaining the cohesive cohesin rings. Disruption in the Rad21-SA2 interaction leads to precocious sister chromatid separation. We proposed a model for how SA2 maintains the cohesin ring. In the model, SA2 fortifies the cohesin ring through providing the tension for the stabilized ring structure (Figure 5.7). The disruption of the Rad21-SA2 interaction leads to the dissociation of SA2 from the cohesin complex and resulting in the collapse and opening of the cohesin ring/s. This model can also be used to explain how cohesin complexes dissociate from the sister chromatids. In prophase, when

SA2 is phosphorylated by Plk1, the phosphorylation can cause the conformational change of the SA2 molecule, which can lead to the open of the cohesin ring/s in a few possible ways: 1) by elongating the SA2 molecule, the tension of which, will break the ring; 2) by shortening the SA2 molecule, which leads to the loss of the tension and the collapse of the ring; 3) by interfering with the interaction of SA2 and the unknown protein, leading to the collapse of the ring. After the ring is open, the cohesin complex is dissociated from the sister chromatids. How SA2 strengthens and stabilizes the cohesin ring is an important question and our present results provide a framework for future studies.

6.3.2 The interaction of SA2 with other proteins

A recent study has showed that SA2 (162-290aa) is responsible for interacting with the transcription regulator CTCF (Xiao *et al.*, 2011), consistent with our finding that N-terminal of SA2 (1-302aa) is not critical in the interaction with Rad21. We have identified that the SA2 (301-750aa) is important for interacting with Rad21 and the C-terminus of SA2 (1052-1231aa) is dispensable for the Rad21-SA2 interaction. In metazoans, phosphorylation of SA2 by Polo-like kinase 1(Plk1) at the chromosome arms during prophase is required for dissolution of arm cohesion (Sumara *et al.*, 2000; Sumara *et al.*, 2002; Gimenez-Abian *et al.*, 2004). The C-terminus of SA2 contains 12 of the 14 SA2 phosphorylation sites, including the 7 highly phosphorylated sites (Hauf *et al.*, 2005). Therefore, this region may be structurally flexible and exposed for the phosphorylation by Plk1.

Despite the phosphorylation sites at the C-terminus, the interaction sites with Rad21 in the middle and CTCF at the N-terminus (Xiao *et al.*, 2011), SA2 may also associate with other cohesin-related proteins. Previous studies showed that the yeast Scc3 and Rad61 can directly form complexes *in vitro* (Rowland *et al.*, 2009). Therefore, it is possible that the human SA2 and Pds5A/B can directly interact with each other. However, this interaction region of SA2 is yet to be determined.

6.3.3 Comparison of the human SA1 and SA2

SA1, the other Scc3 ortholog in humans, shares ~70% sequence identify with SA2. Most of the amino acid variations between SA1 and SA2 are in the 1-68aa and 1075-1162aa regions. SA2 (301-751aa) shares 77% sequence identity with SA1, suggesting similar molecular interactions between SA1 and Rad21. Our result showed that the Rad21 mutants incapable of interacting with SA2 were also unable to interact with SA1 (Figure 5.4), thus confirming the above postulation. It is also in consistence with the finding that the Rad21 362-404aa is crucial for interacting with SA1 (Shintomi *et al.*, 2009). Our finding has narrowed down the interaction region between Rad21 and SA1.

We have shown that SA1 and SA2 interact with Rad21 in similar manners, but it is unknown how they function differently in cells. It is likely that their association with other proteins leads to the differentiated function. The variation in their association with other proteins is likely a result of the sequence variation in the N- and C- termini, which are much less conservative than other regions.

6.4 Future directions

There are goals I would like to pursue in the future for this project: structural studies of Rad21 and SA2 and characterization of the unknown protein participating in the formation of a cohesive cohesin ring. By pursuing these goals, we will have a better understanding of the protein-protein interaction of the human cohesin complex and their physiological function.

6.4.1 Structural studies

We have characterized that both SA2 and Rad21 contain disordered regions which made them difficult to crystallize. We have developed a few new schemes based upon our results from structural characterization and functional studies of the two proteins.

We have proposed new deletion mutant candidates of Rad21 and SA2 for crystallization studies. In addition to the crystallization method, AU and CD methods will be used to provide information on the shape and secondary structure of the molecules. NMR will be used as an alternative method for solving the structure of Rad21 and SA2.

In addition to the structural studies of a single protein, we have also proposed to study the structure of a complex. We have identified 10 amino acids in Rad21 that are critical for interacting with SA2. This 10aa peptide can be used to form complex with an SA2 deletion mutant. For SA2, we have identified amino acids 301-750 are important.

However, we have not narrowed down this region. More constructs with progressive 50 amino acids increment/decrement from the N- and C-terminus of the 301-750 region will be made using the mammalian cell system to further characterize the interaction between Rad21 and SA2. Once a smaller region is identified, we will use the insect cell expression system to express the identified region. Meanwhile, the 10 amino acid peptide of Rad21 will be synthesized using a peptide synthesis system to obtain large amount of peptide. The purified SA2 region and the 10aa Rad21 peptide will be mixed *in vitro* to form complex. The mixture will be purified using a gel filtration column. We should expect a left-shift of the peak of the complex comparing to the chromatogram of SA2 alone. MALDI-MS can also be used to determine if the 10aa peptide is in the complex. If it exists in the complex, we expect to see a peak at the MW of ~1.25kD. The purified complex will then be used for crystallography and NMR studies.

6.4.2 Characterization of the unknown protein involved in the cohesin structure maintenance

It was mentioned in section 5.3 that SA2 might associate with Smc3 through the binding to an unknown protein. This unknown protein can be Pds5A/B, Wapl or another unidentified protein.

To test if Pds5A/B and/or Wapl play the role of the unknown protein, immunoprecipitation products of Rad21 from the mitosis arrested cells ectopically expressing WT or mutant Rad21 will be examined. If the amount of immunoprecipitated Pds5 or Wapl is reduced in the cells expressing mutant Rad21, it is very likely that they

play an important role in the maintenance of the cohesin ring. If neither Pds5A/B nor Wapl shows any difference in the immunoprecipitation, it might indicate that there is another protein involved in this process. To look for this protein, the IP products will be analyzed using in gel digestion and Mass Spec. The proteins detected by MS in both WT Rad21 and mutant Rad21 IP will be compared to each other. If there is a protein present in the IP product of WT Rad21 but not in that of the mutant Rad21, this is protein may be the unknown protein we are searching for.

Bibliography

Anderson, D. E., Losada, A., Erickson, H. P., and Hirano, T. (2002). Condensin and cohesin display different arm conformations with characteristic hinge angles. J Cell Biol *156*, 419-424.

Arumugam, P., Gruber, S., Tanaka, K., Haering, C., H., Mechtler, K., and Nasmyth, K. (2003). ATP hydrolysis is required for cohesin's association with chromosomes. Curr Biol *13,* 1941-1953.

Arumugam, P., Nishino, T., Haering, C., H., Gruber, S., Nasmyth, K. (2006). Curr Biol. *16,* 1998-2008.

Bondos, S. E., and Bicknell, A. (2003). Detection and prevention of protein aggregation before, during and after purification. Anal Biochem *316*, 223-31.

Brookes, E., Cao, W., and Demeler, B. (2010). A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. Eur Biophys J *39,* 405-414.

Canudas, S., and Smith, S. (2009). Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. J Cell Biol *187,* 165-173.

Cao, W., and Demeler, B. (2008) Modeling analytical ultracentrifugation experiments with an adaptive space-time finite element solution for multicomponent reacting systems. Biophys J *95*, 54-65.

Choma, D., Daures, J. P., Quantin, X., and Pujol, J. L. (2001). Aneuploidy and prognosis of non-small-cell lung cancer: a meta-analysis of published data. Br J Cancer *85*, 14-22.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol Cell *5,* 243-254.

Demeler, B. (2005). Ultrascan: a comprehensive data analysis software package for analytical ultracentrifugation experiments. In Modern Analytical Ultracentrifugation: Techniques and Methods, D.J. Scott, S.E. Harding, and A.J. Rowe, eds. (London: Royal Society of Chemistry), pp. 210-229.

Demeler, B. (2010). Methods for the design and analysis of sedimentation velocity and sedimentation equilibrium experiments with proteins. Current protco protein sci Chapter 4, unit 7.13.

Demeler, B., and van Holde, K., E. (2004). Sedimentation velocity analysis of highly heterogenous systems. Anal Biochem *335,* 279-288.

Duesberg, P., Rausch, C., Rasnick, D., and Hehlmann, R. (1998). Genetic instability of cancer cells is proportional to their degree of aneuploidy. Proc Natl Acad Sci U S A *95*, 13692-13697.

Dwawiche, N., Freeman, L., A., and Strunnikov, A., (1999). Characterization of the components of the putative mammalian sister chromatid cohesion complex. Gene *233*, 39-47.

Furth, P.A. (1999). Apoptosis and the development of breast cancer. In "Breast Cancer: Molecular Genetics, Pathogenesis, and Therapeutics" (Eds. A. M. Bowcock and N.J. Totowa), Humana Press, NY, p171-180.

Ghiselli, G., and Lozzo, R., V. (2000). Overexpression of bamacan/SMC3 causes transformation. J Biol Chem *275*, 20235-20238.

Gimenez-Abian, J., F., Sumara, I., Hirota, T., Hauf, S., Gerlich., D., de la Torre, C., Ellenberg, J., and Peters, J., M. (2004). Regulation of sister chromatid cohesion between chromosome arms. Curr Biol *14*, 1187-1193.

Gruber, S., Arumugam, P., Katou, Y., Kuglitsch, D., Helmhart, W., Shirahige, K., and Nasmyth, K. (2006). Evidence that loading of cohesin onto chromosomes involves opening of its SMC hinge. Cell *127*, 523-537.

Gruber, S., Haering, C. H., and Nasmyth, K. (2003). Chromosomal cohesin forms a ring. Cell *112*, 765-777.

Guacci, V., Koshland, D., and Strunnikov., A. (1997). A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. Cell *91*, 47-57.

Guu, T., S., Dong, L., Wittung-Stafshede, P., and Tao, Y., J. (2008). Mapping the domain structure of the influenza A virus polymerase acidic protein (PA) and its interaction with the basic protein 1 (PB1) subunit. Virology *379*, 135-142.

Haering, C., H., Loewe, J., Hochwagen, A., and Nasmyth, K. (2002). Molecular architecture of SMC proteins and the yeast cohesin complex. Mol Cell *9*, 773-788.

Haering, C. H. and Nasmyth, K. (2003). Building and breaking bridges between sister chromatids. Bioessays *25*, 1178-1191.

Haering, C.H., Schoffnegger, D., Nishino, T., Helmhart, W., Nasmyth, K., and Lőwe J. (2004). Structure and Stability of Cohesin's Smc1-Kleisin Interaction. Mol Cell *15*, 951-964.

Hagstrom, K. A., and Meyer, B. J. (2003). Condensin and cohesin: more than chromosome compactor and glue. Nature Reviews Genetics *4*, 520-534.

Hall, T., A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series *41*, 95-98.

Hartman, T., Stead, K., Koshland, D., and Guacci, V. (2000). Pds5p is an essential chromosomal protein required for both sister chromatid cohesion and condensation in Saccharomyces cerevisiae. J Cell Biol *151,* 613-626.

Hanisch, A., Wehner, A., Nigg, E., A., and Sillje., H., H. (2006).Different Plk1 functions show distinct dependencies on Polo-Box domain mediated targeting. Mol. Biol. Cell *17,* 448-459.

Hauf, S., Roitinger, E., Koch, B., Dittrich, C., M., Mechtler, K., and Peters, J., M. (2005). Dissociation of cohesin from chromosome arms and loss of arm cohesion during early mitosis depends on phosphorylation of SA2. Plos Biol *3,* e69.

Hauf, S., Waizenegger, I. C., and Peters, J. M. (2001). Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science *293,* 1320-1323.

Hirano, T. (2002). The ABCs of SMC proteins: two-armed ATPases for chromosome condensation, cohesin, and repair. Genes Dev *16,* 399-414.

Huang, J. and Moazed, D. (2006). Sister chromatids cohesion in silent chromatin: each sister to her own ring. Genes Dev *20,* 132-137.

Ikeguchi, M., Ohfuji, S., Oka, A., Tsujitani, S., Maeta, M., Kaibara, N. (1995). Aneuploidy of tumor cells in cases of gastric cancer with esophageal invasion: another indicator of poor prognosis. J Surg Oncol *58,* 83-90.

Kitajima, T., S., Sakuno, T., Ishiguro, K., Iemura, S., Natsume, T., Kawashima, S., A., and Watanabe, Y. (2006). Shugoshin collaborates with protein phosphatase 2A to protect cohesin. Nature *441,* 46-52.

Kraft, C., Herzog, F., Gieffers, C., Mechtler, K., Hagting, A., Pines, J., Peters, J. M. (2003). Mitotic regulation of the human anaphase-promoting complex by phosphorylation. EMBO J *22,* 6598-6609.

Kueng, S., Hegemann, B., Peters, B., H., Lipp, J., J., Schleiffer, A., Mechtler, K., and Peters., J., M. (2006). Wapl controls the dynamic association of cohesin with chromatin. Cell *127,* 955-967.

Kurze, A., Michie, K., A., Dixon, S., E., Mishra, A., Itoh, T., Khalid, S., Strmechi, L., Shirahige, K., Haering, C., H., Lowe, J. et al. (2011). A positively charged channel within the Smc1/Smc3 hinge required for sister chromatid cohesion. EMBO J *30,* 364-378.

Linding, R., Russell, R., B., Neduva, V., Gibson, T., J. (2003). GlobPlot: exploring protein sequence for globularity and disorder. Nucl Acids Res *31,* 3701-3708.

Losada, A., Yokochi, T., Kobayashi, R., and Hirano, T. (2000). Identification and characterization of SA/Scc3p subunits in the Xenopus and Human cohesin complexes. J Cell Biol *150(3),* 405-416

Losada, A., Yokochi, T., and Hirano, T. (2005). Functional contribution of Pds5 to cohesin mediated cohesion in human cells and Xenopus egg extracts. J Cell Sci *118,* 2133-2141.

Lőwe, J., Cordell, S.C., and van den Ent, F. (2001). Crystal structure of the SMC protein complexes required for sister chromatid cohesion. Genes Dev *12,* 1986-1997.

McGuinness, B., E., Hirota, T., Kudo, N., R., Peters, J. M., Nasmyth, K. (2005). Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. Plos Biol *e86,* 433-449.

Melby, T., E., Ciampaghio, C., N., Briscoe, G., and Erickson, H., P. (1998). The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. J Cell Biol *142,* 1595-1604.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91,* 35-45.

Milutinovich, M. and Koshland, D. E. (2003). Molecular biology. SMC complexes-wrapped up in controversy. Science *300,* 1101-1102.

Nacopoulou, L., Panayotopoulou, E., G., Giannopoulou, I., Tsirmpa, I., Katsarou, S., Mylona, E., Alexandrou, P., Keramopoulos, A. (2007). Extra copies of chromosomes 16 and X in invasive breast carcinomas are related to aggressive phenotype and poor prognosis. J Clin Pathol *60,* 808-815.

Nasmyth, K. (2001). Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. Annu Rev Genet *35,* 673-645.

Nasmyth, K., and Haering, C., H. (2009). Cohesin: its roles and mechanisms. Annu Rev Genet *43,* 525-558.

Nishiyama, T., Ladurner, R., Schmitz, J., Kreidl, E., Schleiffer, A., Bhaskara, V., Bando, M., Shirahige, K., Hyman, A., A., Mechtler, K., et al. (2010). Sororin mediates sister chromatid cohesion by antagonizing Wapl. Cell *143,* 737-749.

Panigrahi, A., K., Pati, D. (2009). Road to the crossroads of life and death: linking sister chromatid cohesion and separation to aneuploidy, apoptosis and cancer. Crit Rev Oncol Hematol *72,* 181-193.

Panizza, S., Tanaka, T., Hochwagen, A., Eisenhaber, F., Nasmyth, K. (2000). Pds5 cooperates with cohesin in maintaining sister chromatid cohesion. Curr Biol *10,* 1557-1564.

Pati, D., Zhang, N., Plon, S., E. (2002). Linking sister chromatid cohesion and apoptosis: role of Rad21. Mol Cell Biol *22,* 8267-8277.

Peters, J. M. (2002). The anaphase-promoting complex: proteolysis in mitosis and beyond. Mol Cell *9,* 931-943.

Peters, J., M., and Bhaskara, V. (2009). Cohesin acetylation: from antiestablishment to establishment. Mol Cell *34,* 1-2.

Peters, J., M., Tedeschi, A., and Schmitz., J. (2008). The cohesin complex and its roles in chromosome biology. Genes Dev *22,* 3089-3114.

Pie, J., Gil-rodriguez, M., C., Ciero, M., Lopez-Vinas, E., Ribate, M., P., Arnedo, M., Deardorff, M., A., Puisac, B., Legarreta, J., de Karam, J., C., et al. (2010). Mutations and variants in the cohesion factor genes NIPBL, SMC1A, and SMC3 in a cohort of 30 unrelated patients with Cornelia de Lange syndrome. Am J Med Genet A *152,* 924-929.

Rankin, S., Ayad, N., G., and Kirschner, M., W. (2005) Sororin, a substrate of the anaphase-promoting complex, is required for sister chromatid cohesion in vertebrates. Mol Cell *18,* 185-200.

Rasband, W., S. ImageJ, U.S. National Institute of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/, 1997-2008.

Rost, B., Yachdav, G., and Liu, J. (2003). The PredictProtein Server. Nucl Acid Res *32,* W321-W326.

Rowland, B., D., Roig, M., B., Nishini, T., Kurze, A., Uluocak, P., Mishra, A., Beckouet F., Underwood, P., Metson, J., Imre, R., et al. (2009). Building sister chromatid cohesion: smc3 acetylation counteracts an antiestablishment activity. Mol Cell *33,* 763-774.

Rubio, E., D., Reiss, D., J., Welcsh, P., L., Disteche, C., M., Filippova, G., N., Baliga, N., S., Aebersold, R., Ranish, J., A., and Krumm, A. (2008). CTCF physically links cohesin to chromatin. Proc Natl Acad Sci U S A *105,* 8309-8314.

Sciallero, S., Giaretti, W., Geido, E., Bonelli, L., Zhankui, L., Saccomanno, S., Zeraschi, E., and Pugliese, V. (1993). DNA aneuploidy is an independent factor of poor prognosis in pancreatic and peripancreatic cancer. Int J Pancreatol *14,* 21-28.

Schmitz, J., Watrin, E., Lenart, P., Mechtler, K., and Peter, J., M. (2007). Sororin is required for stable binding of cohesion to chromatin and for sister chromatid cohesion in interphase. Curr Biol *17,* 630-636.

Shintomi, K., and Hirano, T. (2009). Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. Genes Dev *23,* 2224-2236.

Sumara, I., Vorlaufer, E., Gieffers, C., Peters, B., H., and Peters, J., M. (2000). Characterization of vertebrate cohesin complexes and their regulation in prophase. J Cell Biol *151,* 749-762.

Sumara, I., Vorlaufer, E., Stukenberg, P., T., Kelm, O., Redemann, N., and Peters, J., M. (2002). The dissociation of cohesin from chromosomes in prophase is regulated by Polo-like kinase. Mol Cell. *9,* 515-525.

Sun, Y., Kucej, M., Fan, H., Y., Yu, H., Sun, Q., Y., and Zou, H. (2009). Separase is recruited to mitotic chromosomes to dissolve sister chromatid cohesion in a DNA-dependent manner. Cell *137,* 123-132.

Schmitz, J., Watrin, E., Lenart , P., Mechtler, K., and Peters, J., M. (2007) Sororin is required for stable binding of cohesin to chromatin and for sister chromatid cohesion in interphase. Curr Biol *17*, 630-636.

Terret, M., E., Sherwood, R., Rahman, S., Qin, J., and Jallepalli, P., V. (2009). Cohesin acetylation speeds the replication fork. Nature *462*, 231-234.

Uhlmann, F. (2001). Chromosome cohesion and segregation in mitosis and meiosis. Curr Opin Cell Biol *13,* 754-761.

Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999). Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1. Nature *400,* 37-42.

Uhlmann, F., Wernic, D., Poupart, M. A., Koonin, E. V., and Nasmyth, K. (2000). Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. Cell *103,* 375-386.

Unal, E., Heidinger-Pauli, J., M., Kim, W., Guacci, V., Onn, V., Onn, I., Gygi, S., P., and Koshland, D., E. (2008). A molecular determinant for the establishment of sister chromatid cohesion. Science *321,* 566-569.

Vass, S., Cotterill, S., Valdeolmillos, A. M., Barbero, J. L., Lin, E., Warren, W., D., and Heck, M., M. (2003). Depletion of Drad21/Scc1 in Drosophilia cells leads to instability of the cohesin complex and disruption of mitotic progression. Curr Biol *13*, 208-218.

Viadiu, H., Stemmann, O., Kirschner, M., W., Walz, T. (2005). Domain structure of separase and its binding to securin as determined by EM. Nat Struct Mol Biol *12*, 552-553.

Waizenegger, I., C., Hauf, S., Meinke, A., Peter, J., M. (2000). Two distinct pathways remove mammalian cohesin from chromosome arms in prophase and from centromeres in anaphase. Cell *103,* 399-410.

Xiao, T., Wallace, J., and Felsenfeld, G. (2011). Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex, and are required for cohesin dependent insulation activity. Mol Cell Biol *31*, 2174-2183.

Yeh, E. Haase, J., Paliulis, L., V., Joglekar, A., Bond, L., Bouck, D., Salmon, E., D., and Bloom, K., S. (2008). Pericentric chromatin is organized into an intramolecular loop in mitosis. Curr Biol *18*, 81-90.

Zhang, N., Ge, G., Meyer, R., Sethi, S., Basu, D., Pradnan, S., Zhao, Y., J., Li, X., N., Cai., W., W., El-Naggar, L., et al. (2008). Overexpression of Separase induces aneuploidy and mammary tumorigenesis. Proc Natl Acad Sci U S A *105*, 13033-13038.

Zhang, N., Kuznetsov, S., G., Sharan, S., K., Li, K., Rao, P., and Pati, D. (2008). A handcuff model for cohesin complex. J Cell Biol *183(6):* 1019-1031.

Zidovetzki, R., Rost, B., Armstrong, D., L., and Pecht, I. (2003) .Transmembrane domains in the functions of Fc receptors. Biophys Chem *100,* 555-575.

Zou, H., McGrarry, T., J., Bernal, T., Kirschner, M., W. (1999). Identification of a vertebrate sister-chromatid separation inhibitor involved in transformation and tumorigenesis. Science *285,* 418-422.

Appendices

Appendix I Recipes for buffers

Hypotonic lysis buffer (1x): 50 mM Tris-HCl, pH=7.5, 50mM NaCl, 1% Triton X-100 and 1mM PMSF

Phosphate-buffered saline (PBS, 1x): 137mM NaCl, 2.7mM KCl, 8.1mM Na_2HPO_4 and 1.5mM KH_2PO_4, pH=7.5

Tris-buffered saline (TBS, 1x): TBS, 50mM Tris-HCl, 150mM NaCl, pH=7.5

FPLC buffer A: 50mM Tris-HCl, pH=7.5, 1 mM EDTA, 2mM βME, and 1 mM NaN_3

Gel filtration buffer: 50mM Tris-HCl, 200mM NaCl, 10% Glycerol, 1mM EDTA, 2mM β-mE, and 1mM NaN_3

Ni-NTA Buffer A: 50mM Tris-HCl, pH 7.5, 300mM NaCl, 20mM imidazole, 10% glycerol, and 0.1mM PMSF

Ni-NTA Buffer B: 50mM Tris-HCl, pH 7.5, 300mM NaCl, 250mM imidazole, 10% glycerol, and 0.1mM PMSF

Ni-NTA wash buffer (denaturative condition): 50mM Tris-HCl, 8M Urea, pH=6.3

Ni-NTA elution buffer 1: 50mM Tris-HCl, 8M Urea, pH=5.9

Ni-NTA elution buffer 2: 50mM Tris-HCl, 8M Urea, pH=4.5

Refolding buffer: 50mM Tris-HCl, 20mM NaCl, 0.8mM KCl, pH=8.2

Dialysis Buffer: 50mM Tris-HCl, 20% Glycerol, 500mM NaCl, and 1mM EDTA

SDS sample buffer (2x): 125mM Tris-HCl, pH=6.8, 20% glycerol, 4% SDS, 0.1% bromophenol blue, and 100mM DTT

TBST buffer: pH=7.4, 20mM Tris-HCl, 150m NaCl, and 0.1% Tween 20

Gel filtration buffer (for AU sample preparation): 25mM Tris-HCl pH=7.5, 200mM NaCl, and 5% glycerol

Lysis buffer (for mammalian cells): 50mM Tris-HCl, pH=7.4, 1% Triton X-100, 150mM NaCl, 1mM EDTA, Complete mini Protease Inhibitor tablet (Roche; 1 tablet in every 10ml lysis buffer)

Appendix II Protein sequences

A 2.1 Cohesin subunit SA2 [Homo sapiens] (GI: 112789530)

MIAAPEIPTDFNLLQESETHFSSDTDFEDIEGKNQKQGKGKTCKKGKKGPAEKGKGGNGGGKPPSGPNRM
NGHHQQNGVENMMLFEVVKMGKSAMQSVVDDWIESYKHDRDIALLDLINFFIQCSGCKGVVTAEMFRHMQ
NSEIIRKMTEEFDEDSGDYPLTMAGPQWKKFKSSFCEFIGVLVRQCQYSIIYDEYMMDTVISLLTGLSDS
QVRAFRHTSTLAAMKLMTALVNVALNLSINMDNTQRQYEAERNKMIGKRANERLELLLQKRKELQENQDE
IENMMNAIFKGVFVHRYRDAIAEIRAICIEEIGIWMKMYSDAFLNDSYLKYVGWTMHDKQGEVRLKCLTA
LQGLYYNKELNSKLELFTSRFKDRIVSMTLDKEYDVAVQAIKLLTLVLQSSEEVLTAEDCENVYHLVYSA
HRPVAVAAGEFLYKKLFSRRDPEEDGMMKRRGRQGPNANLVKTLVFFFLESELHEHAAYLVDSMWDCATE
LLKDWECMNSLLLEEPLSGEEALTDRQESALIEIMLCTIRQAAECHPPVGRGTGKRVLTAKEKKTQLDDR
TKITELFAVALPQLLAKYSVDAEKVTNLLQLPQYFDLEIYTTGRLEKHLDALLRQIRNIVEKHTDTDVLE
ACSKTYHALCNEEFTIFNRVDISRSQLIDELADKFNRLLEDFLQEGEEPDEDDAYQVLSTLKRITAFHNA
HDLSKWDLFACNYKLLKTGIENGDMPEQIVIHALQCTHYVILWQLAKITESSSTKEDLLRLKKQMRVFCQ
ICQHYLTNVNTTVKEQAFTILCDILMIFSHQIMSGGRDMLEPLVYTPDSSLQSELLSFILDHVFIEQDDD
NNSADGQQEDEASKIEALHKRRNLLAAFCKLIVYTVVEMNTAADIFKQYMKYYNDYGDIIKETMSKTRQI
DKIQCAKTLILSLQQLFNEMIQENGYNFDRSSSTFSGIKELARRFALTFGLDQLKTREAIAMLHKDGIEF
AFKEPNPQGESHPPLNLAFLDILSEFSSKLLRQDKRTVYVYLEKFMTFQMSLRREDVWLPLMSYRNSLLA
GGDDDTMSVISGISSRGSTVRSKKSKPSTGKRKVVEGMQLSLTEESSSSDSMWLSREQTLHTPVMMQTPQ
LTSTIMREPKRLRPEDSFMSVYPMQTEHHQTPLDYNRRGTSLMEDDEEPIVEDVMMSSEGRIEDLNEGMD
FDTMDIDLPPSKNRRERTELKPDFFDPASIMDESVLGVSMF

A2.2 Cohesin subunit Rad21 [Homo sapiens] (GI: 5453994)

MFYAHFVLSKRGPLAKIWLAAHWDKKLTKAHVFECNLESSVESIISPKVKMALRTSGHLLLGVVRIYHRK
AKYLLADCNEAFIKIKMAFRPGVVDLPEENREAAYNAITLPEEFHDFDQPLPDLDDIDVAQQFSLNQSRV
EEITMREEVGNISILQENDFGDFGMDDREIMREGSAFEDDDMLVSTTTSNLLLESEQSTSNLNEKINHLE
YEDQYKDDNFGEGNDGGILDDKLISNNDGGIFDDPPALSEAGVMLPEQPAHDDMDEDDNVSMGGPDSPDS
VDPVEPMPTMTDQTTLVPNEEEAFALEPIDITVKETKAKRKRKLIVDSVKELDSKTIRAQLSDYSDIVTT
LDLAPPTKKLMMWKETGGVEKLFSLPAQPLWNNRLLKLFTRCLTPLVPEDLRKRRKGGEADNLDEFLKEF
ENPEVPREDQQQQHQQRDVIDEPIIEEPSRLQESVMEASRTNIDESAMPPPPPQGVKRKAGQIDPEPVMP
PQQVEQMEIPPVELPPEEPPNICQLIPELELLPEKEKEKEKEKEDDEEEEDEDASGGDQDQEERRWNKRT
QQMLHGLQRALAKTGAESISLLELCRNTNRKQAAAKFYSFLVLKKQQAIELTQEEPYSDIIATPGPRFHI
I

Appendix III List of Clones

Vector	Insert	N-terminal Tag
pFastBac	SA2	HA
	SA2	His
	SA2 (1-150)	His
	SA2 (1-302)	His
	SA2 (1-450)	His
	SA2 (1-587)	His
	SA2 (1-750)	His
	SA2 (1-903)	His
	SA2 (1-1051)	His
	SA2 (1-1122)	His
	SA2 (151-1051)	His
	SA2 (303-1051)	His
	SA2 (451-1051)	His
	SA2 (581-1051)	His
	SA2 (751-1051)	His
	SA2 (904-1051)	His
	SA2 (1052-1231)	His
pET-28b	Rad21 (451-631)	His
	Rad21 (280-450)	His
pFastBac	Rad21	Flag
	Rad21 (1-172)	Flag
	Rad21 (1-450)	Flag
	Rad21 (171-631)	Flag
	Rad21 (451-631)	Flag
	Rad21 (171-450)	Flag
	Rad21 (171-350)	Flag
	Rad21 (171-382)	Flag
	Rad21 (171-417)	Flag
	Rad21 (198-450)	Flag
	Rad21 (205-450)	Flag
	Rad21 (240-450)	Flag
	Rad21 (280-450)	Flag
	Rad21 (316-450)	Flag
	Rad21 (351-450)	Flag
	Rad21 (383-450)	Flag
	Rad21 (171-450)	His

pFlagCMV2	Rad21 (171-350)	Flag
Vector	**Insert**	**N-terminal Tag**
pFlagCMV2	Rad21 (171-364)	Flag
	Rad21 (171-382)	**Flag**
	Rad21 (171-392)	Flag
	Rad21 (171-400)	**Flag**
	Rad21 (171-417)	Flag
	Rad21 (171-450)	**Flag**
	Rad21 (351-450)	Flag
	Rad21 (364-450)	**Flag**
	Rad21 (393-631)	Flag
	Rad21 (401-631)	**Flag**
	Rad21 (418-631)	Flag
	Rad21 (436-631)	**Flag**
	Rad21 (L385A)	Flag
	Rad21 (F389A)	**Flag**
	Rad21 (L383A)	Flag
	Rad21 (V397A)	**Flag**
	Rad21 (L385AF389A)	Flag
	Rad21 (L385AT390A)	**Flag**
	Rad21 (F389AT389A)	Flag
	Rad21 (L385AF389AT389A)	**Flag**
	Rad21 del (383-392)	Flag
	Rad21 del (393-400)	**Flag**
	Rad21 del (383-400)	Flag
pCS2MT	SA2 (1-1051)	Myc
	Rad21 (L385AF389A)	**Myc**
	Rad21 (L385AT390A)	Myc
	Rad21 (F389AT389A)	**Myc**
	Rad21 (L385AF389AT389A)	Myc
	Rad21 del (383-392)	**Myc**